HARVEY MOTULSKY

Biostatistica essenziale

Una guida non matematica

Edizione italiana a cura di
Leonardo Emberti Gialloreti
Dipartimento di Biomedicina e Prevenzione
Università degli Studi di Roma "Tor Vergata"

Traduzione di
Daniele di Giovanni
Dipartimento di Ingegneria Industriale
Università degli Studi di Roma "Tor Vergata"

Con la collaborazione di
Fabio Riccardi
Annalisa Aragozzini

PICCIN

This is a translation of
ESSENTIAL BIOSTATISTICS: A Nonmathematical Approach, First Edition
by Harvey Motulsky

Essential Biostatistics: A Nonmathematical Approach, First Edition was originally published in English in 2016. This translation is published by arrangement with Oxford University Press. Piccin Nuova Libraria is solely responsible for this translation from the original work and Oxford University Press shall have no liability for any errors, omissions or inaccuracies or ambiguities in such translation or for any losses caused by reliance thereon.

L'edizione originale in lingua inglese di *Essential Biostatistics: A Nonmathematical Approach*, First Edition è stata pubblicata nel 2016. Questa traduzione è stata pubblicata su licenza di Oxford University Press. Piccin Nuova Libraria è responsabile della traduzione dall'opera originale e Oxford University Press non è responsabile per eventuali errori, omissioni, sviste o ambiguità della traduzione o per eventuali danni da essa derivanti.

ISBN 978-88-299-3162-0

Stampato in Italia

www.piccin.it

INDICE GENERALE

PRESENTAZIONE ALL'EDIZIONE ITALIANA

Comprendere a fondo un lavoro scientifico e fare ricerca è impossibile senza una qualche nozione di statistica. Oggi fare un'analisi statistica è molto più facile che nel passato per la semplicità di utilizzo dei molti software in commercio. L'accessibilità a questi strumenti rende più semplice, per tutti, affrontare l'analisi dei dati ma di pari passo aumenta il rischio di commettere errori e di giungere a conclusioni sbagliate. Questo accade molto più spesso di quanto si pensi.

Si può *imparare* la statistica su molti eccellenti testi. Il volume che avete in mano è invece pensato per chi vuole *imparare a utilizzare* la statistica. Imparare e imparare ad utilizzare: posso dire, dopo tanti anni di insegnamento universitario, che non è una differenza da poco. Molti infatti possono fare un calcolo statistico, ma non tutti sono in grado di affrontare una situazione reale che richiede di applicare un'analisi statistica e di comprenderne il senso.

È per questo motivo che volentieri ho accolto l'invito dell'Editore Piccin di curare *Biostatistica essenziale: una guida non matematica*, edizione italiana del rinomato *Essential Biostatistics: A Nonmathematical Approach* di Harvey Motulsky, un testo che è progettato per chi vuole concretamente utilizzare la statistica in prima persona, ma che è pensato anche per chi desidera comprendere il reale messaggio che i dati statistici possono trasmettere.

È un testo da utilizzare come compagno di viaggio, in parallelo o quale approfondimento ad un corso universitario di statistica sanitaria. Non si sofferma su tutti i dettagli (qualche purista forse storcerà il naso, pazienza) ma accompagna il lettore a *fare statistica* e ad accorgersi in tempo dei possibili *errori* in cui è facile incorrere.

Speriamo di offrire alla comunità universitaria e ai ricercatori italiani un agevole *sussidio* per scegliere con competenza le tecniche statistiche più utili a rispondere a domande scientifiche reali e per comprendere e utilizzare consapevolmente la terminologia statistica.

Prof. Leonardo Emberti Gialloreti

PREFAZIONE

Biostatistica essenziale: una guida non matematica è un sussidio non matematico, conciso ed accessibile, al pensiero statistico. È concepita come un breve testo di accompagnamento ai corsi di statistica universitaria e alla lettura di libri di statistica più lunghi che adottano un approccio più matematico o anche come strumento di revisione per i ricercatori.

In che modo questa edizione è diversa dall'edizione completa di *Intuitive Biostatistics*?

Intuitive Biostatistics venne pubblicato per la prima volta nel 1995, ed è ora alla sua quarta edizione. Tuttavia, sebbene in molti abbiano elogiato il libro, alcuni docenti lo trovavano poco adatto al loro corso perché non comprendeva tutti quei contenuti matematici che ci si attenderebbe di trovare in un libro di testo principale; d'altra parte, risultava anche troppo lungo per essere utilizzato come testo complementare. Per colmare questa lacuna, ho scritto *Biostatistica essenziale: una guida non matematica*, lungo circa un terzo di *Intuitive Biostatistics* (3ª edizione).

Alcuni tra gli argomenti presenti in *Intuitive Biostatistics* ma assenti da *Biostatistica essenziale: una guida non matematica* sono la meta-analisi, le curve di sopravvivenza, i test di equivalenza o di non inferiorità, la regressione non lineare e multipla, il confronto della bontà di adattamento tra modelli alternativi e o le curve ROC. Inoltre, *Intuitive Biostatistics* (ma non *Biostatistica essenziale: una guida non matematica*) comprendeva anche più di 100 pagine di spiegazione dei test statistici di base, nonché 40 pagine di esercizi con risposte.

Un approccio unico

Ecco alcuni aspetti che rendono questo libro unico nel suo genere:

- Il Capitolo 1 spiega come il senso comune possa portare a conclusioni errate e perché è necessario comprendere i principi della statistica.
- Il Capitolo 2 propone un approccio unico per rendersi conto di quanto sia complesso il concetto di probabilità.

- Introduco il pensiero statistico a partire dal Capitolo 4, dove si spiega l'intervallo di confidenza per la proporzione. Posso così spiegare la logica alla base dell'utilizzo dell'intervallo di confidenza per generalizzare da un campione ad una popolazione, prima di dover affrontare il concetto della dispersione. Non commettete l'errore di trascurare il Capitolo 4 solo perché pensate di non dover lavorare con dati espressi come delle proporzioni, poiché questo capitolo spiega alcuni dei più importanti concetti della statistica.
- Introduco il confronto tra gruppi tramite gli intervalli di confidenza (Capitolo 12) prima di spiegare il *p-value* (Capitolo 13) e la significatività statistica (Capitoli 14 e 15). In questo modo posticipo la presentazione dei concetti disorientanti e fin troppo abusati di *p-value* e di "significativo".
- Il Capitolo 16 spiega quanto sono comuni gli errori di tipo I e qual è la differenza tra significatività e *false discovery rate*.
- Il Capitolo 19 presenta, sotto forma di tabelle, i più comuni test statistici.
- Ho incluso alcuni argomenti spesso omessi dai testi introduttivi, ma che considero essenziali. Ad esempio, i confronti multipli, il *false discovery rate*, il "dragaggio dei dati" (*p-hacking*), le distribuzioni log-normali, la media geometrica, i test di normalità, gli *outlier* o la regressione non lineare.
- Quasi ogni capitolo ha una sezione dedicata alla *terminologia*, dove si spiegano alcuni possibili fraintendimenti linguistici.
- Per aiutarvi ad evitare interpretazioni errate dei dati, quasi ogni capitolo include anche una sezione che discute gli *errori più frequenti*. L'intero Capitolo 25 è, infine, incentrato sui principali errori da evitare.

Chi ha aiutato?

Un immenso grazie alle molte persone che hanno rivisto le bozze dei capitoli. I loro contributi hanno enormemente migliorato questo libro:

Abdel-Salam G. Abdel-Salam, Virginia Polytechnic Institute and State University
B. Carol Adjemian, Pepperdine University
Raid Amin, University of West Florida
Michael Biro, Swarthmore College
Patrick Breheny, University of Kentucky
Michael F. Cassidy, Marymount University
Dean W. Coble, Stephen F. Austin State University
William M. Cook, Saint Cloud State University
Erica A. Corbett, Southeastern Oklahoma State University
Vincent A. DeBari, Seton Hall University
Bianca Frogner, George Washington University
Robert M. Hamer, University of North Carolina at Chapel Hill
Philip Hejduk, University of Texas at Arlington

Ravi P. Joshi, Old Dominion University
Stefan Judex, Stony Brook University
Chris Kerth, Texas A&M University
Joshua Lallaman, Saint Mary's University of Minnesota
Gary A. Lamberti, University of Notre Dame
Bruce D. Leopold, Mississippi State University
Susan P. McGorray, University of Florida
Matt McQueen, University of Colorado Boulder
Sumona Mondal, Clarkson University
Christopher J. Salice, Texas Tech University
David A. Sanchez, University of Texas at Arlington
Evelyn H. Schlenker, University of South Dakota
Andrew Jay Tierman, Saginaw Valley State University
Kathryn Trinkaus, Washington University
Derek Webb, Bemidji State University
Mary M. Whiteside, University of Texas at Arlington
John W. Wilson, University of Pittsburgh
Naji Younes, George Washington University

Ringrazio anche tutte le persone della Oxford University Press che hanno contribuito a trasformare il mio manoscritto in un libro raffinato: Jason Noe, Senior Editor; Andrew Heaton, Assistant Editor; Patrick Lynch, Editorial Director; John Challice, Publisher and Vice President; Bill Marting, National Sales Manager; Frank Mortimer, Director of Marketing; David Jurman, Marketing Manager; Elizabeth Geist, Marketing Assistant; Lisa Grzan, Production Manager; Amy Whitmer, Production Team Leader; Christian Holdener, Senior Production Manager; Michele Laseau, Art Director; e Bonni Leon-Berman, Designer.

La mia esperienza in materia

Per oltre un decennio ho insegnato statistica agli studenti del primo anno di medicina e ai laureati in scienze biomediche dell'Università della California, San Diego. Il programma di questi corsi si è ampliato fino a divenire la prima edizione di *Intuitive Biostatistics*, il testo originale e completo su cui si basa questo libro. Non insegno più in questi corsi, ma in qualità di fondatore e CEO di GraphPad Software, scambio e-mail con studenti e ricercatori quasi ogni giorno. Resto quindi costantemente aggiornato sui molti modi in cui i concetti statistici possono confondere o essere fraintesi.

Scrivetemi per consigli, correzioni o suggerimenti. Eventuali correzioni saranno riportate nel sito www.intuitivebiostatistics.com.

Harvey Motulsky
hmotulsky@graphpad.com