

Prefazione

I biologi moderni hanno bisogno dei metodi statistici per analizzare i dati che raccolgono. È per questo motivo che un numero sempre crescente di università offre agli studenti di biologia, o persino richiede, un corso di base sull'analisi dei dati. Negli ultimi vent'anni abbiamo tenuto un corso di questo tipo alla University of British Columbia, e durante tutti questi anni abbiamo cercato un testo che trattasse gli argomenti nel modo più adatto a un corso propedeutico. Quasi tutti i libri, però, erano troppo tecnici ed enciclopedici, oppure non erano abbastanza completi, perché privi di metodi cruciali per la pratica della biologia moderna. Desideravamo un testo che privilegiasse un tipo di comprensione intuitiva dei concetti invece di fare troppo affidamento sulle formule. E volevamo insegnare l'analisi dei dati attraverso esempi interessanti. Ma soprattutto, avevamo bisogno di un libro con un forte taglio biologico, che affrontasse gli argomenti importanti su cui si interrogano i biologi quando raccolgono e analizzano i dati reali.

Non siamo riusciti a trovare il testo che volevamo e quindi abbiamo deciso di scriverne uno per colmare la lacuna. Crediamo che questo libro abbia molte caratteristiche che lo distinguono da altri testi; caratteristiche che, come abbiamo scoperto, sono molto utili per raggiungere efficacemente il nostro pubblico.

Esempi biologici interessanti. L'esperienza dell'insegnamento ci ha dimostrato che gli studenti di biologia imparano meglio l'analisi dei dati nel contesto di esempi interessanti tratti dalla letteratura medica e biologica. La statistica è un mezzo per raggiungere un fine, uno strumento per conoscere la natura. Enfatizzando tutto ciò che possiamo apprendere in ambito biologico, la forza e l'utilità della statistica diventano evidenti. E la statistica diventa anche più divertente se applicata a ricerche interessanti.

Ogni capitolo utilizza diversi esempi per presentare i concetti chiave e ogni esempio è descritto nel suo contesto biologico. Gli esempi sono corredati da immagini degli organismi reali, e quindi gli studenti possono guardare ciò che stanno imparando. L'importanza data allo studio di esempi reali e interessanti si ritrova anche nell'abbondanza degli esercizi proposti: per ogni capitolo sono presentate decine di problemi basati su dati reali che riguardano argomenti di ambito biologico.

Spiegazioni intuitive dei concetti chiave. Il ragionamento statistico richiede nuovi modi di pensare. Ma gli studenti rischiano di perdersi, bombardati da molti termini nuovi e innumerevoli test. Ci siamo resi conto, però, che è estremamente utile partire sempre dall'intuizione, tralasciando i dettagli. Abbiamo quindi adot-

tato un approccio intuitivo per rispondere alle domande fondamentali. Che cos'è un buon campione? Che cos'è un intervallo di confidenza? Perché facciamo un esperimento? I primi capitoli introducono queste conoscenze di base, quelli successivi le sviluppano.

Analisi dei dati in pratica. Questo libro si concentra sui dati piuttosto che sulle basi matematiche della statistica. Insegnamo come si fanno buone visualizzazioni grafiche ed evidenziamo il fatto che un diagramma è il punto di partenza di ogni buona analisi dei dati. Dedicamo lo stesso tempo al processo di stima e alla verifica delle ipotesi, ed evitiamo di considerare il *P*-value come fine a sé stesso. Il libro non richiede una conoscenza della matematica al di là dell'algebra più semplice. Ci concentriamo sugli aspetti pratici piuttosto che sulle sfumature, sull'utilità della statistica in biologia invece che sui problemi teorici. Insegnamo non soltanto il modo «giusto» di fare qualcosa, ma poniamo in rilievo anche alcuni dei tranelli che si potrebbero incontrare.

Assumiamo che per la maggior parte dei calcoli sarà disponibile un computer, e quindi ci concentriamo sui concetti dell'analisi dei dati biologici e su come la statistica ci può aiutare a capire meglio la natura. Grazie ai moderni computer che abbiamo a disposizione, oggi la parte fondamentale nell'analisi dei dati è capire quale metodo usare e perché.¹ Immaginiamo e speriamo che chi usa questo libro abbia anche la possibilità di fare pratica con i software statistici. Siamo anche consapevoli che la diversità di questi programmi è enorme e quindi non abbiamo legato il libro ad alcun particolare pacchetto software.

Disegno sperimentale in pratica. Un biologo non può applicare la statistica in modo corretto – o in generale non può fare buona scienza – senza una comprensione pratica del disegno sperimentale. A differenza della maggior parte dei testi, qui esaminiamo anche gli argomenti fondamentali nel disegno sperimentale, quali i controlli, la randomizzazione, la pseudo-replicazione e il raggruppamento (blocking), e lo facciamo in modo intuitivo e considerando gli aspetti pratici.

Concetti di base aggiornati. Che lo crediate o no, il migliore intervallo di confidenza di una proporzione non

¹ «Un computer vi permette di commettere errori a una velocità superiore a quella di qualsiasi altra invenzione nella storia dell'uomo – con le possibili eccezioni della pistola e della tequila.» Mitch Ratcliffe, in *Technology Review*, 1992.

è quello che forse avete imparato in passato. E le statistiche non parametriche non valutano efficacemente eventuali differenze tra medie (o tra mediane) senza fare alcune assunzioni piuttosto forti, che solitamente non vengono considerate. In questo testo abbiamo aggiornato questi e molti altri argomenti di base della statistica.

Trattazione di argomenti moderni. La biologia moderna utilizza un insieme di strumenti statistici più ampio di quello in uso una generazione fa. In questo libro ci spingiamo oltre i temi trattati dalla maggior parte dei testi propedeutici, discutendo anche i principi concettuali di argomenti importanti come la verosimiglianza, la regressione non lineare, la randomizzazione, la metanalisi e il bootstrap.

Utili riepiloghi. I concetti chiave sono sempre presentati in un breve riepilogo alla fine di ogni capitolo. La maggior parte dei capitoli termina con un paragrafo di riepilogo delle formule, che raggruppa la maggior parte delle equazioni facilitandone la consultazione.

Schede. Tra un capitolo e l'altro sono presentate brevi trattazioni di temi specifici che abbiamo chiamato «schede». Le schede considerano un'ampia varietà di argomenti concettuali e di senso comune che sono cruciali per l'interpretazione dei risultati statistici nella ricerca scientifica. Alcune schede sono focalizzate sui modi in cui la scienza può sbagliare e su come tenere conto di questi errori. Anche se le schede non sono comprese nei capitoli, integrano la materia trattata nei capitoli principali. Consigliamo caldamente di leggerle.

Avete tra le mani il risultato di cinque anni di lavoro. Crediamo che *L'analisi statistica dei dati biologici*, trattando un'ampia gamma di argomenti in modo pratico e intuitivo, fornisca un buon background sull'analisi dei dati che ogni biologo dovrebbe avere. È molto apprezzato nei nostri corsi; speriamo che lo sia anche nei vostri.

L'organizzazione del libro

L'analisi statistica dei dati biologici è suddiviso in cinque parti, ciascuna formata da un certo numero di capitoli. Consigliamo di iniziare la lettura dalla prima parte, perché essa introduce molti concetti fondamentali poi utilizzati in tutto il libro. Questi primi capitoli dovrebbero essere letti nella loro interezza.

Dopo la prima parte, molti dei capitoli successivi procedono presentando gli argomenti più generali introdotti all'inizio e passando ad argomenti più specifici verso la fine. Ogni capitolo è strutturato in modo che i primi paragrafi permettano di ottenere una comprensione fondamentale dell'argomento. Per esempio, nel capitolo sull'analisi della varianza (Capitolo 15) i concetti fondamentali sono presentati nei primi due paragrafi; i Paragrafi 15.1 e 15.2 presentano circa gli stessi concetti che si possono trovare nella maggior parte dei testi propedeutici di statistica. I Paragrafi 15.3-15.6 trattano aspetti addizionali e altre applicazioni interessanti.

L'ultima parte del libro (Capitoli 18-21) si rivolge principalmente al lettore «avventuroso» e curioso. Questi capitoli introducono alcuni argomenti, quali la verosimiglianza, il bootstrap e la metanalisi, che si incontrano comunemente nella letteratura medica e biologica, ma che spesso non vengono menzionati in un corso propedeutico. Introducono i concetti fondamentali di ciascun argomento, spiegano come funzionano i metodi e indicano dove reperire ulteriori informazioni.

Un corso di base potrebbe comprendere soltanto i Capitoli 1-17, senza andare oltre i Paragrafi 5.6, 7.3, 8.4, 9.3, 12.6, 13.6, 15.2, 16.4 e 17.5. Sugeriamo che tutti i corsi pongano in risalto gli argomenti trattati nelle schede.

Ogni capitolo termina con una serie di problemi ed esercizi che si propongono di valutare la comprensione, da parte degli studenti, dei concetti e delle applicazioni pratiche della statistica. I problemi sono suddivisi in «problemi di approfondimento» e «ulteriori problemi». Brevi risposte ai problemi di approfondimento sono fornite alla fine del libro.

Una parola sui dati

I dati usati in questo libro sono reali, con alcune eccezioni ben evidenziate. La maggior parte di questi dati è stata presa direttamente da articoli pubblicati. In alcuni casi abbiamo contattato direttamente gli autori degli articoli che ci hanno fornito generosamente i dati grezzi. Spesso, quando questi non erano presenti nell'articolo originale, abbiamo deciso di risalire ai dati digitalizzando i grafici originali, come i diagrammi di dispersione e gli istogrammi. Inevitabilmente i valori numerici che abbiamo estratto in questo modo differiscono lievemente da quelli originali. In rari casi abbiamo generato al computer dati che concordassero con le analisi statistiche presentate nell'articolo. I risultati che presentiamo sono sempre compatibili con le conclusioni degli articoli originali.

Ringraziamenti

Se fossimo stati solo in due, non saremmo mai riusciti a pubblicare questo libro. Molte altre persone hanno fornito contributi rilevanti. La chiarezza e l'accuratezza del contenuto del libro sono state migliorate dalla meticolosa attenzione di un gran numero di lettori, tra cui Arianne Albert, Brad Anholt, Cecile Ane, Eric Baack, James Bryant, Martin Buntinas, C. Ray Chandler, Christiana Drake, Jonathan Dushoff, Steven George, Aleeza Gerstein, George Gilchrist, Brett Goodwin, Mike Hickerson, Darren Irwin, Nusrat Jahan, Philip Johns, Istvan Karsai, Robert Keen, John Kelly, Rex Kenner, Ben Kerr, Joseph G. Kunkel, Todd Livdahl, Brian C. McCarthy, Eli Minkoff, Robert Montgomerie, Spencer Muse, Courtney Murren, Claudia Neuhauer, Patrick C. Phillips, Jay Pitochelli, James Robinson, Simon Robson, Michael Rosenberg, Noah Rosenberg, Nathan Rank, Bruce Rannala, Mark Rizzardi, Michael Russel, Ronald W. Russel, Andrew Schaffner, Andrea Schluter, William Thomas, Michael Travisano, Thomas Valone, Bruce Walsh, Grace A. Wyngaard e Sam Yeaman. Molte di queste persone hanno letto più capitoli e tutte hanno contribuito a migliorare la chiarezza e l'accuratezza del libro. Sally Otto e Allan Stewart-Oaten si sono meritati la nostra eterna gratitudine leggendo l'intero libro e comunicandoci le loro osservazioni. Ovviamente, siamo solo noi i responsabili di eventuali errori; non sempre abbiamo accolto i consigli che ci sono stati dati, anche quando forse avremmo dovuto farlo. Se abbiamo dimenticato il nome di qualcuno, lo ringraziamo.

Siamo debitori verso gli studenti di BIOL 300 della University of British Columbia, che hanno sperimentato in classe questo libro nel corso degli ultimi anni. Il testo ha beneficiato anche dell'adozione in parecchi college e università, nei corsi tenuti da Brad Anholt (University of Victoria), Eric Baack (Luther College), George Gilchrist (College of William and Mary), Mike Hickerson (Queens College, City University of New York), Nusrat Jahan (James Madison University), Susan Lehman (Brock University), Jean Richardson (Brock University), Simon Robson (James Cook University) e Grace A. Wyngaard (James Madison University). George Gilchrist e i suoi studenti ci hanno fornito una serie molto particolareggiata ed estremamente utile di osservazioni in uno stadio cruciale del libro. I seguenti studenti della University of British Columbia (UBC) e di altre istituzioni hanno scoperto errori rilevanti nelle bozze del libro: Jessica Beaubier, Edward Cheung, Lorena Cheung, Stephanie Cheung, Denise Choi, Samrad Ghavimi, Inderjit Grewal, Sarah Hamanishi, Gurpreet Khaira, Jung Min Kim, Arleigh Lambert, Alexander Leung, Mira Li, Flora Liu, Dianna Luouie, Johnston Mak, Sarah Neumann, Ruth Ogbamichael, Marion Pearson, Trevor Schofield, Meredith Soon, Erin Stacey, Michelle Uzelac, Hillary Ward, Chris Wong, Irene Yu, Anush Zakaryan, Paul Zhou e Jon-Paul Zacharias.

Numerosi ricercatori ci hanno inviato cortesemente i loro dati originali, tra cui Matt Arnegard, Andrey Bar-

ker-Plotkin, Butch Brodie, Pamela Colosimo, Kevin Fowler, Chris Harley, Luke Harmon, Andrew Hendry, Peter Keightley, Fredrik Liljeros, Jean Thierry-Mieg, Jeffrey S. Mogil, Patrik Nosil, Margarita Ramos, Rick Relyea, Jake Socha, Brian Starzomski, Richard Svanback, Andrew Trites, Jason Weir, Jack Werren e Martin Wikelski.

Il libro è stato sottoposto a editing e revisione dallo straordinario editor John Murdzek. John ha lavorato oltre il normale orario di lavoro per produrre questo libro, con un'agenda zeppa di impegni, trovando persino un refuso in «*Photinus ignites*». Tom Webster ha dato una piacevolissima veste grafica ai diagrammi e alle illustrazioni, Laura Roberts ha dedicato molte ore a cercare le meravigliose fotografie e Jeff Whitlock ha generosamente fornito molte delle belle fotografie che compaiono in questo libro. Mark Ong e la sua équipe presso gli Side by Side Studios hanno compiuto il lavoro fantastico di trasformare il manoscritto in un libro. Siamo particolarmente grati a Eric Baack per aver lavorato duramente sugli esercizi. Gunder Hefta e Aleeza Gerstein hanno scovato numerosi errori con la loro accurata correzione di bozze. Infine, Ben Roberts merita i nostri più vivi ringraziamenti per il sostegno che ci ha dato nella realizzazione di questo libro.

La stesura del libro è iniziata quando Michael Whitlock lavorava presso il Peter Wall Institute for Advanced Studies (UBC) in qualità di Distinguished-Scholar-in-Residence, e la maggior parte delle parti finali del testo è stata scritta mentre era Sabbatical Scholar presso la National Evolutionary Synthesis Center nella North Carolina. Dolph Schluter ha cominciato a lavorare al libro mentre era visiting professor in Biologia dello sviluppo presso la Stanford University. Il sostegno e l'ambiente accademico forniti da ciascuna di queste istituzioni sono stati eccezionali, e a esse siamo molto grati.

Infine, vogliamo ringraziare tutte le persone che ci hanno insegnato molto nel corso degli anni. Michael Whitlock desidera ringraziare Dave McCauley, Mike Wade, Nick Barton, Ben Pierce, Kewin Fowler, Patrick Phillips, Sally Otto e Betty Whitlock.

Gli autori

Michael Whitlock è biologo evolucionista e genetista di popolazioni. È professore di zoologia alla University of British Columbia, dove insegna statistica agli studenti di biologia dal 1995. Attualmente è Editor-in-Chief di «*The American Naturalist*».

Dolph Schluter è professore e Canada Research Chair nello Zoology Department and Biodiversity Research Center presso la University of British Columbia. È noto per le sue ricerche sull'ecologia e sull'evoluzione dei fringuelli delle Galápagos e dello spinarello. È membro delle Royal Societies of Canada and London.

Prefazione all'edizione italiana

L'analisi statistica dei dati è indispensabile per studiare e capire gli organismi viventi nel loro habitat naturale e per interpretare i risultati della sperimentazione. L'enorme variabilità biologica che ci circonda deve essere analizzata con un metodo quantitativo e oggettivo che permetta di giungere a conclusioni che non siano solo «impressioni», o «idee», ma che possano veramente convincerci che un'ipotesi è (o non è) da preferire rispetto a un'altra, o che ci aiutino a stimare una caratteristica che osserviamo e a definire la precisione della stima.

L'importanza dell'analisi statistica è riconosciuta in quasi tutti i corsi di laurea che si tengono in Italia nell'ambito delle scienze della vita. Sotto varie denominazioni, come Biostatistica o Metodi Statistici, ogni sede universitaria offre insegnamenti o moduli propedeutici (e a volte anche avanzati) sull'impiego della statistica per analizzare i dati biologici. Purtroppo, però, gli studenti iscritti a questi corsi di laurea trovano ancora qualche difficoltà nell'apprendere la statistica di base, e a volte mostrano una certa riluttanza a formulare ragionamenti statistico-matematici, applicare formule o a comprendere i passaggi logici delle dimostrazioni. Sono i docenti che devono fare di più, cercando di integrare efficacemente l'analisi dei dati con il problema biologico, la logica del metodo statistico con il suo utilizzo pratico nel contesto della disciplina, coinvolgendo gli studenti con esempi reali e interessanti. Non bisogna altresì dimenticare che la biologia moderna e l'analisi dei dati biologici sono molto cambiate negli ultimi anni, anche grazie alla diffusione del personal computer.

Il libro di Whitlock e Schluter va esattamente incontro a queste esigenze. Quando la Casa editrice Zanichelli mi ha chiesto gentilmente notizie del testo che mi ero impegnato a scrivere già diversi anni fa, e anche un parere sulla possibile traduzione in italiano del libro di Whitlock e Schluter, ho avuto un'illuminazione. Questo è il testo che avevo in mente di scrivere, ma non riuscivo mai a trovare il tempo di andare oltre i primi capitoli introduttivi. Si capisce fin dalle prime pagine che gli autori hanno una grande esperienza sia come biologi sia come insegnanti di statistica in corsi di laurea di area biologica. Il risultato è eccellente: un testo in cui si amalgamano efficacemente i dati biologici nel loro contesto e le metodologie statistiche, quasi «tradotte»

per chi è interessato soprattutto ai problemi biologici.

Il libro contiene innanzitutto il materiale di base necessario per gli studenti di molti corsi di laurea triennali in ambito biologico e naturalistico, ma anche per esempio medico e agrario. Include poi molti argomenti più avanzati adatti alle lauree magistrali o a corsi post-laurea. Ed è anche un testo che sarà di grandissima utilità per colleghi biologi e ricercatori, che troveranno molte tecniche statistiche utili per i loro studi, descritte e spiegate in modi nuovi, semplici, chiari, e corredate da efficaci esempi tratti dalla letteratura scientifica. Gli ultimi capitoli del libro, dedicati alle metodologie più moderne, sono fondamentali per rendere l'opera completa e adatta non solo a studenti ma anche a laureati e colleghi che giornalmente lavorano analizzando i dati biologici che hanno faticosamente raccolto.

Ringrazio i dott. Giorgia Valpiani, Silvia Fuselli, Livio Finos e Andrea Adanese che mi hanno aiutato nella supervisione della traduzione. Non è stato un compito facile, la statistica ha il suo gergo e la terminologia è a volte ambigua e non univoca. Molti termini inglesi hanno due, tre o anche quattro possibili traduzioni, e a volte nessuna di queste è incisiva come la parola inglese (che quindi non ho tradotto). In alcuni rari casi ho anche deciso di utilizzare indifferentemente termini sinonimi per abituare il lettore a familiarizzare con essi. Mi scuso per eventuali scelte inappropriate. Il mio intento è stato quello di rendere la trattazione più semplice possibile senza appesantimenti terminologici, e allo stesso tempo di utilizzare i termini più comunemente impiegati nella letteratura scientifica e nei software per l'analisi statistica. Ho anche cercato di mantenere lo stile molto informale, a volte ironico e divertente, della versione inglese. Questo, insieme a piccole modifiche concordate con gli autori, può avere generato delle differenze rispetto al testo originale. Me ne assumo ovviamente la responsabilità, ma sono comunque convinto che l'edizione italiana abbia mantenuto la chiarezza dell'originale.

Giorgio Bertorelle
Professore di Biostatistica
e di Biologia Evoluzionistica
Università di Ferrara