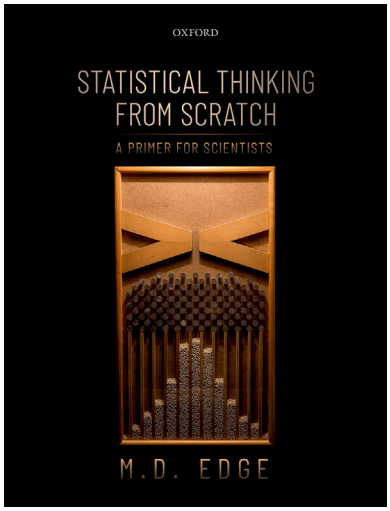
Please cite this article in press as: Guzman and Pennell, Mathematical Statistics for Biologists and Other Interesting People, Trends in Ecology & Evolution (2019), https://doi.org/10.1016/j.tree.2019.09.005

Book Review

Mathematical Statistics for Biologists and Other Interesting People

Laura Melissa Guzman[1,*] and Matthew W. Pennell[2,*]



As ecologists and evolutionary biologists, along with many of their colleagues across the sciences, have become increasingly statistically sophisticated, there has been a deluge of books, blogs, and tweets explaining how to make sense of data. Many of these are great (many are not). However, most of the resources aimed at nonstatisticians skirt around the mathematical reasoning that underlies procedures; that is to say they often focus on the how of statistics and not the why. *Statistical Thinking from Scratch: A Primer for Scientists* [1], a new book by M.D. Edge, a population geneticist, fills a unique niche in this landscape, sitting between the inference-focused material most biodiversity scientists are likely familiar with, and mathematical statistics books that focus on the derivations and properties of estimators. Edge's conceit is that any scientist with a basic familiarity of algebra and calculus – and some serious motivation – can gain a rich understanding of where inference methods come from and why they work (and also when they might not). Rather than work through a variety of different methodologies, almost the entire book is devoted to explaining a single technique, bivariate linear regression, at a deep level; the big concepts can then be ported to other operations.

This is a rather ambitious objective and overall, we think Edge succeeds spectacularly. Edge guides readers step-by-step through derivations of summary statistics and estimators, accompanying the mathematical notation with intuitive explanations. And where he suspects readers could use a refresher on some background concept, he takes the space to walk them through it like a patient teacher. We both teach statistics in various capacities, and consequently, often found ourselves in awe of how clear Edge's explanations of seemingly familiar topics were; we will certainly be borrowing liberally from this book in the future. We honestly enjoyed reading the book and thinking through the concepts and arguments, and suspect that many other scientists will as well.

Alongside the exposition in the text, *Statistical Thinking from Scratch* includes an abundance of exercises, spread throughout the book. We found these to be a particular highlight. They are (well) designed to help readers build their statistical intuition using a variety of techniques, including analyses of simple datasets, simulation, and mathematical proofs. All of the exercises from the book have detailed solutions, either in the back of the book or on the associated GitHub repository [2]. Edge has also written a comprehensive and exceptionally easy-to-use R package that contains all the necessary functions for running the simulations described in the exercises. While these functions serve as an excellent starting point, we suspect that in some cases it would be worthwhile for readers to write their own simulators to better understand what is going on under the hood.

Although the book does a truly fantastic job explaining concepts, it is rather light on examples, and all of these are based on rather silly premises ('are people in the state of Arizona shorter, on average, than the rest of Americans?') or rely on simple, made-up datasets (Anscombe's famous quartet [3] transmogrifies into data on African crop yields). We are not sure if more or better empirical examples would have made the mathematical concepts clearer, but in our own teaching, we have found that working through real-world problems helps the students understand the objectives of inference methods better. The book is titled *Statistical Thinking from Scratch*, but we anticipate that readers who are not familiar with how statistics are used 'in the wild' may struggle to grasp the point of it all. Similarly, we think that the introduction to R is probably too brief for a true novice. As such, we are of the opinion that the book is probably too advanced for an introductory biostatistics course for undergraduates – we teach using Whitlock and Schluter's *The Analysis of Biological Data* [4] and will continue to do so (full disclosure: both of us have, at some time,

Please cite this article in press as: Guzman and Pennell, Mathematical Statistics for Biologists and Other Interesting People, Trends in Ecology & Evolution (2019), https://doi.org/10.1016/j.tree.2019.09.005

shared a hallway with Whitlock and Schluter) – but would unhesitatingly recommend it to curious scientists of all types and career stages. This is a truly unique field guide to the statistical jungle.

Statistical Thinking from Scratch: A Primer for Scientists by M.D. Edge, Oxford University Press, 2019. 9780198827627, US$85.00/£65.00

[1]Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

[2]Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

*Correspondence:
laura_melissa_guzman@sfu.ca (L.M.G.),
pennell@zoology.ubc.ca (M.W.P.)

https://doi.org/10.1016/j.tree.2019.09.005

References

1. Edge, M.D. (2019) *Statistical Thinking from Scratch: A Primer for Scientists* (Oxford University Press)
2. Edge, M.D. (2018) Code and supplemental material for *Statistical Thinking from Scratch*. *GitHub*. https://github.com/mdedge/stfs
3. Anscombe, F.J. (1973) Graphs in statistical analysis. *Am. Stat.* 27, 17–21
4. Whitlock, M.C. and Schluter, D. (2015) *The Analysis of Biological Data*, 2nd edn (Macmillan Learning)