Book Review

Simplicity and serenity in advanced statistics

Modern Statistics of the Life Sciences by Alan Grafen and Rosie Hails. Oxford University Press, 2002. £22.99 pbk (384 pages) ISBN 0 19 925231 9. **Statistical Computing. An Introduction to Data Analysis Using S-Plus** by Michael J. Crawley. John Wiley & Sons, 2002. US$85.00 hbk (772 pages) ISBN 0 471 56040 5

Andrew F. Read

Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK
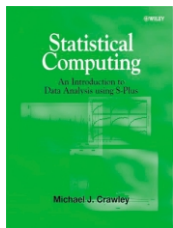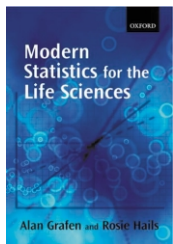
One of the great intellectual triumphs of the 20th century was the discovery of the generalized linear model (GLM). This provides a single elegant and very powerful framework in which 90% of data analysis can be done. Conceptual unification should make teaching much easier. But, at least in biology, the textbook writers have been slow to get rid of the historical baggage. These two books are a huge leap forward.

Modern Statistics of the Life Sciences is a translation of lecture handouts that Grafen provided for his Quantitative Methods Course at Oxford University. These handouts, which have been circling the globe for more than ten years, provoked very different reactions among those forced to do the course and those who sought them out. Fortunately for Oxford undergraduates, the book is a huge improvement. It is the first coherent exposition of statistical model building using general linear models that is accessible for near-novices. It assumes some basic knowledge (to the level of Samuels [1], say), but then, uniquely in a market place packed with texts, it develops the whole unified philosophy in a coherent way from scratch for biologists. It would be great to teach from [fun even, especially given the free Powerpoint presentations (http:// www.oup.com/grafen-hails)], but it is also an excellent self-teach guide for graduate students and beyond. It deals only with general linear models, but after a thorough study of this book, students will be doing sophisticated analyses and be well placed to attack generalized linear models.

The clarity of the material in the book varies from exemplary to average. Chapter 1 is the best exposition I have read on the basics of ANOVA. [This is the sample chapter available free (http://www.oup.com/grafenhails). If you are ANOVA-allergic, read this and see just how easy it really is]. The discussion of the principles of model choice and model checking are also excellent. But for those of us unable to offer the luxury of 1:1 tutorials, the geometric view is a distraction and, for the 2nd edition, the sections about nesting and random effects need reworking. Moreover, students need persuading that violations of assumptions can lead to mistaken conclusions, and that rectifying them is not fiddling the data. Underwood [2] does an excellent job of this.

When learning statistics, the computer can interfere with understanding. Grafen and Hails get around this by being program independent. Their book provides no information about how to get a computer to do anything. It studies generic output. Comprehensive package specific supplements (Minitab, SPSS, SAS) are provided free (http://www.oup.com/grafenhails), as are all the data sets, but the book stands alone. This makes the book accessible to anyone, no matter which software is flavour of the month. Crawley's book is different. He wants you to learn GLM through S + (or the freeware analogue R), because of its enormous generality and superior graphics. And he's right: if you want to learn just one package, this is it. But as an over view of the philosophy of modern analysis, S + seems to get in the way, unlike the program GLIM in Crawley's earlier masterpiece [3]. Beginners unable to study the book on Crawley's annual course might be better starting with *Modern Statistics of the Life Sciences*.

Nonetheless, *Statistical Computing* is surely going to be the key point of access for biologists wanting the raw grunt of S + . The book has some peerless sections, such as the introductions to bootstrapping and jack knifing, the diagnosis and remedy of poorly fitting models, and the sections about survival analysis and repeated measures. There are even extra free chapters (http://www.bio.ic.ac.uk/research/mjcraw/statcomp/). Best, though, is the Crawley enthusiasm and wit [after an analysis of a split−split−split−split (!) plot experiment: 'Interpretation of output tables like this requires a high level of serenity'].

Both books make it clear that much of the power of the GLM approach comes from model comparison: seek simplicity and then distrust it. The pedagogical challenge is to nurture sensible trust/distrust decisions. For instance, Crawley argues that if we take significance as $P < 0.05$ and draw conclusions from 40 P-values, then at least two of them will be false positives. Grafen and Hails point out that the problem is much worse: if there is a 95% chance of avoiding a false positive, then, in 40 tests, there is a 88% chance (0.95^{40}) of at least one false positive. Even in a simple experiment, the figure is nasty. A factorial experiment with three variables will have seven P-values in the ANOVA table (three main effects, four interactions) − a 30% chance of a false positive.

This issue is widely appreciated, and contentious [4]. Even estimating the scale of the problem is difficult. The

---

calculations above would almost certainly overestimate the false positive rate in the literature. They assume, for instance, that nothing is going on, that the tests are independent, and that a hypothesis lives or dies by P-value alone. None of this is true in the real world that scientists inhabit. To some extent, the scale of the problem is in the eye of the beholder. Grafen and Hails are clearly worried about it, and they argue, in essence, for less analysis: if you don't peek, you reduce the problem. This is a sound message for beginners, but Crawley teaches us how to peek, recognizing, for instance, that practicing scientists do comparisons between subsets of data.

The bottom line is that statistical analysis does not generate certainty. If we get it right, we get, at best, a good guide to uncertainty and a useful aid to successful inference. That is why statisticians and computers cannot do our job. GLMs provide the very best framework for

discovering the truth in our variable world. But, in spite of their computational sophistication, large dollops of scientific intuition are still required to use them to do good science. I used to think that scientific intuition was common sense, but teaching undergraduates has convinced me otherwise. These books teach the mechanics of GLMs well. But, unlike most other texts, they also teach the vital common sense.

References

1 Samuels, M.L. (1989) *Statistics for the Life Sciences*, MacMillan
2 Underwood, A.J. (1997) *Experiments in Ecology*, Cambridge University Press
3 Crawley, M.J. (1993) *GLIM for Ecologists*, Blackwell
4 Rothmam, K.J. (1990) No adjustments are needed for multiple comparisons. *Epidemiology* 1, 43–46

Plants and animals, forever entangled

Plant–Animal Interactions – An Evolutionary Approach edited by C.M. Herrera and O. Pellmyr. Blackwell Science, 2002. £35.00 pbk (xi + 313 pages) ISBN 0 632 05267 8

Lars Chittka

School of Biological Sciences, Queen Mary, University of London, London E1 4NS, UK
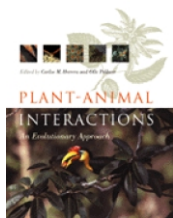
'It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about…and to reflect that these elaborately structured forms…dependent on each other in so complex a manner, have all been produced by laws acting around us,' says Darwin in the final paragraph of *The Origin of Species*. In somewhat more contemporary language, the blurb of *Plant–Animal Interactions* announces 'Attraction, deception, defence, escape and tolerance', but what sounds a bit like the subtitle of a soap opera is in fact referring to the turbulent marriage between animals and plants. Aimed at graduate students, this book shows, however, that studying the intricacies of animal–plant interactions can be substantially more entertaining than those between soap-opera actors. The editors have brought together a team of internationally renowned scholars to write nine chapters about one of the most mature and, simultaneously, most dynamic fields of evolutionary ecology.

Peter W. Price opens the topic with a discussion about the nature and diversity of animal–plant interactions, and how the multiple types of interaction have forged yet more diversity at the species level. He adds an original slant by pointing out that flight is a key innovation in shaping these interactions: most terrestrial organisms – plants or animals – fly at some stage (e.g. plants as pollen or

seeds; and, among the animals, birds, bats and insects, as the most speciose class), and many animal–plant interactions only make sense in the light of flight. Conrad C. Labandeira then presents the rich array of paleobiological and phylogenetic tools used to open windows into the past of the relationship between plants and animals. For example, phylogenetic analyses revealed how insects probably saw colors in the Devonian, long before colored flowers first occurred [1].

There follows a focus on antagonism (insects and mammals as herbivores, phytoparasites and granivores, and plant strategies for deterring the unwanted vegetarians). A convincing point is made that the arms race between herbivores and the defence machinery of affected plants has been the very reason for the extraordinary diversity in all of the groups involved.

The book then turns to mutualism, with the editors each contributing a piece; Olle Pellmyr discussing pollination, and Carlos Herrera, seed dispersal. The relationship between ants and plants, as described by Andrew J. Beattie and Lesley Hughes, is especially intricate and fascinating. Ants pollinate flowers and disperse seeds, they build gardens and protect plants from their foes. In turn, the plants offer the ants various kinds of food and, sometimes, even a home, as exemplified by the ant-acacias of Costa Rica and East Africa.

Finally, John N. Thomson identifies several key avenues for the next generation of scientists interested in the tangled web of plants and animals. Dedicating a whole chapter to the future is an interesting idea, but can we actually predict the course of science? Predictions are

Corresponding author: L. Chittka (l.chittka@qmul.ac.uk).