BOOK REVIEW:

MICHAEL J. CRAWLEY'S STATISTICAL COMPUTING:
AN INTRODUCTION TO DATA ANALYSIS USING S-PLUS

CARSTEN F. DORMANN

Statistical textbooks for ecologists face multiple dilemmas: they are either wide-ranging and shallow or highly specific and profound, they are mathematically exact or more problem-oriented, they are look-up encyclopedias or learn-by-doing course scripts, they use a specific program or they remain vague about the implementation. Building on his success with the earlier book "GLIM for Ecologists" (Blackwell), Crawley's latest statistics book tries to bridge all of these dichotomies. While GLIM was tedious to the extent of being mischievous, S+ is a much neater program, also making the book easier to plough through. The link between the book and the program is a very close one, and users of other statistics packages will have to swap to S+ (or R) to take full profit from the book.

The statement "Nobody ever learned statistics by reading a book about it." (p. ix) certainly holds. Crawley's way of teaching is thus practical. He uses S-plus, a powerful statistical and programming software, which is equipped with high quality graphical output procedures, to work through examples and problems. As he gives the code, the output and the interpretation, every step can be followed easily, although it is sometimes cumbersome to read if you use different software. (Luckily, a "brother" of S+, called "R", can be downloaded free from www.r-project.org and works with the same code in almost all cases. Swapping to S+ or R is well worthwhile, given the versatility and structure of this package.) Thus the reader is guided to a working knowledge in statistics, rather than profound mathematical-statistical knowledge.

Some things about the book are irritating. As statistical theory and worked examples are interwoven, the use as a reference book is difficult. The switches between dry mathematical formulae and Crawley's nonchalant writing style ("prepare to be impressed" (e.g. p. 581); the expression "come into their own" is repeated dozens of times) make an awkward reading, but this is an intrinsic problem of statistical texts. While the index is useful and complete, the table of contents is less than informative.

The book covers a vast range of applied statistical topics, from the obvious introductory methods of ordinary, multiple and non-linear regressions, ANOVAs and GLMs to the advanced problems of different error distributions (Poisson, bionomial, gamma), time series analysis, mixed effect models and spatial statistics. Additionally, Crawley addresses problems important for ecologists such as experimental design, bootstrap & jackknife, split-plot ANOVAs, piecewise regressions and survival analysis. Free internet supplements (http://www.bio.ic.ac.uk/research/mjcraw/statcomp/) add four chapters: gamma errors, nonlinear mixed effect models, general additive models and multivariate statistics, as well as the data sets and all S-plus language code used in the book. A suggested reading at the end of every chapter leads to more specific literature. Two topics neglected by Crawley are non-parametric statistics, as he assumes that most of this can be analysed by parametric statistics with appropriate error distributions and transformations; and Monte Carlo-type re-sampling tests, which he doesn't mention at all.

The downside of touching on every topic is that this approach necessarily remains superficial. For example when discussing bootstrapping, he documents only a standard function, without mentioning the options offered for this procedure. Or the chapter on mixed effect models (with 40 pages one of the longest chapters) still only covers the more simple cases (but then, Pinheiro & Bates devote an entire book to this issue). Crawley's book falls short of teaching the reader S+ beyond the point required for the covered statistical problems, but does a good job of introducing the reader to a panoply of statistical methods. It does not provide a useful overview of S+'s graphical output functions, but is competent with respect to model formulations.

What makes this book very useful is the pragmatic approach Crawley takes. A reknowned ecologists himself, he leads the reader through the jungle of statistical concepts. Crawley clearly distinguishes between important assumptions and (for ecologists) less important mathematical details. He takes the reader through examples that are clearly understandable and are likely to bear a strong resemblance to one's own data. Crawley points out where data collection went wrong, and is blunt about minimal requirements of data (replication, independence). Throughout the book, the reader might feel some detail or special cases are missing, but can rely on the advice given with a tone of conviction.

For anyone giving introductory statistics courses at university level and is willing to do it in S+ (or R), this book provides lots of data, solutions and rules of thumb. Less technical and easier to comprehend than the standard Venables & Ripley (Springer), Crawley's Statistical Computing is a door-opener into S+, and possibly an eye-opener for statistical novices.

Crawley's book is surely not the last statistics book you'll ever need, but it might well be the first you will ever really use.

Carsten F. Dormann
Applied Landscape Ecology
UFZ Leipzig-Halle
Permoserstr.15
04318 Leipzig
Email: dormann@alok.ufz.de