



- 8 Davidian M, Giltinan DM. *Nonlinear models for repeated measurement data*. Boca Raton, FL: Chapman and Hall/CRC, 1998.
- 9 Vonesh EF, Chinchilli VM. *Linear and nonlinear models for the analysis of repeated measurements*. New York: Marcel Dekker, 1997.

Reviewed by Hakan Demirtas, Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, USA.

Altman M, Gill J, McDonald MP. 2004: *Numerical issues in statistical computing for the scientist*. New York: John Wiley & Sons, Inc. 266 pp. \$89.95. ISBN 0 471 23633 0.

Even with an extensive background in statistical theory, a scientist may have many questions remaining about the implementation of statistical procedures on a given set of data. Those with little interest in the exact implementation of the methodology can soon find themselves having to understand implementation details when 'unexpected' issues manifest themselves occasionally through unclear warnings or errors from standard statistical software. Unfortunately, basic understanding of numerical computing as implemented on modern computational hardware is sometimes lacking in even advanced education. This book provides the researcher with an overview of the issues involved in the implementation and computation of common statistical procedures. This should allow the reader to have more confidence in interpreting and troubleshooting common statistical models. The text also gives the reader the ability to better compare the available statistical software and understand the differences that exist.

The text is divided into two sections. The first section is a general overview of numerical computing issues that includes basic problems inherent to numerical computing, methods of evaluating and comparing statistical software, and computational issues that can interfere with statistical inference. The second section presents six specific areas that pose interesting numerical problems considered by both the authors and leading figures in the field of statistics:

- 1) Hessian matrices;
- 2) Markov chain Monte-Carlo estimation;
- 3) King's ecological inference method;
- 4) Nonlinear estimation;
- 5) Spatial Regression Models;
- 6) Convergence in logistic regression.

The first section of the book requires basic knowledge of computer arithmetic and statistical programming as well as intermediate knowledge of statistics that should be possessed by researchers employing inferential methods. The second section of the book is better understood with a background knowledge of specific statistical methodology and application.

The first section presents issues inherent to computational processing that can affect statistical analyses. This section includes a brief explanation of floating point arithmetic, a commonly misunderstood necessity of computing which leads to many questions in both the S-News and R-Help mailing lists. Also included in this section is an overview of random number generation, which is essential to many modern statistical techniques. In addition, an entire chapter in the first section is devoted to the evaluation of statistical software, providing further information on issues that can influence results of statistical analyses. Specific issues discussed in this section include generation and computation of distribution functions, faults in pseudo-random number generators and accuracy of input/output routines. A description of benchmarking using Standard Reference Datasets (StRD) leads into a comparison of statistical packages and languages. I believe that comparisons between statistical packages should be presented in a more real time media because of the frequency of new releases of software that may include bug fixes and updated algorithms. This is especially important when including open-source products such as R. However, information on benchmarking is important so the reader knows how comparisons should be made.

The second section of the text presents the most interesting reading because of its focus on individual statistical techniques. The chapters are well written and provide information in a variety of topics that researchers use when analysing data.

The first area of focus comprises two key components: the importance of the Hessian in likelihood-based estimation, and possible solutions for practitioners of statistical data analysis when facing a non-invertible Hessian. Anyone who has implemented likelihood-based estimation on even an occasional basis has encountered the warning 'Hessian not invertible' from statistical software. Unfortunately the default reaction of many researchers is to change the question they are trying to answer by adjusting the model to avoid getting the warning or error, with no regard to the fact they are changing the original question. Although in some cases this process may not be harmful, it should not be done without careful

consideration and certainly should not be the default method of a researcher. The authors present a solid technical description of the problem of non-invertible Hessians and several alternatives to the usual reaction, including a description of the generalized inverse/generalized Cholesky procedure that they make available through an R package and Gauss procedure.

The chapter devoted to convergence problems in logistic regression, written by Paul Allison, presents information that any practitioner applying this methodology needs to understand. The explanation and impact of complete and quasi-complete separation is clearly written and made easily understandable with simple examples. Warning messages from STATA, Systat and the LOGISTIC procedure from SAS are presented for comparison; output from R, JMP, and the CATMOD and GENMOD procedures from SAS are described. It is also noted that GLIM, LIMDEP, Minitab and the mlogit procedure from STATA give no warnings. Even those readers without an extensive statistical background can readily understand the advice on the reporting of parameter estimates, standard errors and results from the corresponding tests when separation is present. The advice also covers the diagnosis of complete and quasi-complete separation, as well as possible solutions for dealing with these issues. As in the section on the Hessian, the emphasis is on not attempting to delete the variables to make the computation 'work', but rather to focus on the original questions by fitting an appropriate model to answer the pertinent questions as raised from expert knowledge in the given field.

The book is certainly more directed at researchers who are 'users' of statistical software as opposed to those who wish to implement new statistical concepts directly. The text lacks the detail of implementation found in Thisted (1988)¹ or Monahan (2001);² the latter books would better suit a researcher working on implementing methodology not provided by current software. In contrast, Altman, Gill and McDonald have succeeded in providing two useful pieces of information to researchers involved in statistical analysis and interpretation of data: 1) a good understanding of the potential pitfalls involved with the implementation of methodology computationally, and 2) good advice on dealing with the problems that can arise.

References

- 1 Thisted R. *Elements of statistical computing: Numerical computation*. New York, NY: Chapman and Hall, 1988.

- 2 Monahan J. *Numerical methods of statistics*. Cambridge, UK: Cambridge University Press, 2001.

Reviewed by Matthew Austin, Statistician, Amgen, Thousand Oaks, California, USA.

Senn S 2003: *Dicing with death: Chance, risk and health*. Cambridge: Cambridge University Press. 251 + xii pp. £14.95 (PB). ISBN 0 521 54023 2.

Dicing with Death is something of an enigma – it is a book about statistics for the general public. In recent years the market for popular science books has grown quite rapidly, but a notable gap has been for popular statistics books. This may in part be due to the public perception of statistics (or statisticians), but whatever the reason *Dicing with Death* goes some way to redress the balance.

The book tackles a wide range of what to the public are quite difficult problems. However, Senn tackles these with such an elegant manner that even the most complex topics are presented with great clarity. His unified approach is with one idea in mind – that statistics is about the translation of information (data) into knowledge. The book consists of 11 chapters and covers many aspects of health statistics (with a little bit of law for good measure), such as probability and common paradoxes, randomized clinical trials, the Bayesian and frequentist schools of inference, hypothesis/significance testing, life tables and survival analysis, meta-analysis and summarizing evidence, and infectious disease modelling, with the book ending on the somewhat topical subject of the MMR controversy. All throughout the book are an enormous array of quotes, humorous stories and anecdotes, but over and above the 'fun' side is a wealth of information, not least in the historical developments of the ideas that are presented. Underlying all of this is the clear message that statistics is about far more than book-keeping or counting or tabulating numbers, it is a beautiful subject that can help drive scientific research.

Reading the book for this review has been a great pleasure – the enthusiasm of the author for his subject is clear and it rubs off onto the reader. While there are a few equations amongst the text I felt that these were justified and helped nonmathematical readers understand some of the more complex issues. I am sure others would disagree and suggest that any introductory book should be free from mathematical symbols and notation. That aside, the book is suitable for just about any