Journal of Official Statistics
Vol. 1, No. 3, 1985, pp. 351–358

Book Reviews

Books for review are to be sent to the Book Review Editor Arne Sandström, Statistical Research Unit, Statistics Sweden, S–115 81 Stockholm, Sweden.

Bowerman, B. L. and O'Connell, R. T., Computer Modeling for Business and Industry. Statistics: Textbooks and Monographs Series, Vol. 59, Marcel Dekker, Inc., New York and Basel, 1984, ISBN 0-8247-7296-2, vii + 219 pp., SFr. 56.

This book instructs the reader in how to use standard programs in statistics and operations research in four chapters. The methods for analysis are briefly described and illustrated with practical examples. The authors show how to use the SAS package (Statistical Analysis System) for computing descriptive statistics, for hypothesis testing (chi-square and *t*-tests), in simple and multiple regression analysis, and for estimating trend lines. They also show how to decompose a time series by the multiplicative classical decomposition method. This is done by means of a FORTRAN programme called DECOMP, listed in an appendix. The final chapter is an introduction to linear programming using a programme called LINDO.

The aim of the book is to supplement existing texts in statistical and quantitative analysis and integrate the use of computer packages in introductory business statistics and quantitative methods courses. The authors claim that most textbooks do not stress the use of computer packages, even though they may display computer output. This may be true, but there are a few textbooks that try to do more than merely display computer outputs, such as Berenson and Levine (1983) and Gordon and Pressman (1983).

There are many computer packages in these areas that perform about the same work. Differences are primarily found in the manner that data are input and commands given, and not in computer output. A book like this one therefore will be limited to the packages chosen by the authors. For other packages, like the MINITAB or SPSS packages, there are other books that more completely describe different commands, how to use them, and how to interpret the results.

This book is easy to read, with many explained examples. However, the statistical theory behind the methods is not presented, which means that the book needs to be supplemented with a standard textbook on the subject.

In summary, this is a well written book, but there are others serving the same purpose in a more complete fashion.

References

Berenson, M. L. and Levine, D. M. (1983): Basic Business Statistics, Concepts and Applications, Second ed. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Gordon, G. and Pressman, I. (1983): Quantitative Decision-Making for Business, Second ed. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Per-Olov Edlund
Stockholm School of Economics

Farlow, S. J. (Ed.), Self-Organizing Methods in Modelling: GMDH Type Algorithms. Statistics: Textbooks and Monographs, Vol. 54, Marcel Dekker, Inc., New York, 1984, ISBN 0-8247-7161-3, 368 pp., $55.

This book consists of five papers containing general theory (120 pages), eleven papers containing applications (195 pages), and a bibliography. It is devoted to the group method of data handling (GMDH), a modelling method developed by A. G. Ivakhnenko, the Ukrainian cybermitist, and other similar methods.

If a dependent variable y is to be modelled by the full polynomial of a high degree in n basic independent variables $x_1, x_2, \ldots, x_n$, there might be an enormous number of coefficient parameters. The main feature of the GMDH algorithm is that it works stepwise, successively sorting out the best predictors. Predictors are kept or discarded according to a relative root mean square criterion in a separate checking set of data. In each step the new possible predictors are generated as the pairwise (full) second order polynomials in the previous predictors. This means that the separate fitting problems are quite small. The procedure stops when the root mean square criterion reaches its (first!) minimum. In the final fitted model, there might be hundreds of parameters even when there are just a few data points.

"There are no sessions at world or national congresses or conferences devoted to GMDH. Publications in major journals on control or communication theory in the United States and the USSR are not devoted to GMDH papers. This situation must be changed for the benefit of humankind." The words are not mine, but appear in the fifth paper of the book, "Past, Present and Future of GMDH" by A. G. Ivakhnenko. It is typical for some of the authors to exaggerate the importance of GMDH, not to see it as one possible model-building technique, but rather dealing with it as *the* model-building technique. Furthermore, there is a tendency in many places not to realize the unsolved statistical problems related to the method.

Yet, it is a valuable book because it describes the state of the art within GMDH for its theory as well as its application. Currently, the theory consists mainly of a data handling technique, and most of the papers pay no attention at all to the statistical problems that must be considered in order to evaluate the results. In the contribution by Andrew R. Baron (Predicted squared error: A criterion for automatic model selection), however, there are some attempts to treat the randomness inherent in all data.

The application papers come from a great variety of fields. This is exemplified by the following key words from the papers: Agricultural and meteorological time series, river flows, shrimp catches, interest rates, sales amount, economic modelling, environmental system, and management.

A common signature of the applied papers is that they do not use the physical characteristics of the studied subject in the model-building process. The fitted models and the parameters therein have no physical meaning as such, but serve only to determine prediction formulas.

The book is aimed at "modellers in unstructured areas," statisticians, and computer scientists. People with a knowledge of basic mathematical analysis, algebra, and statistics will find it quite easy to read.

But the reader should also be aware of the fact that these methods of data-handling need to be complemented with careful statistical analysis linked to the physical properties of the application problem, in order to achieve full usefulness. Application of the methods without these complements may lead to bad results in several respects.

Sture Holm
University of Göteborg

Efron, B., The Jackknife, the Bootstrap and Other Resampling Plans. CBMS–NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1982, ISBN 0-89871-179-7, vii+92 pp., $12.50.

An important statistical task is to determine the reliability of estimates. During the last ten years, Efron has proposed and studied a new method for computing the standard error called bootstrapping. The method is quite simple to understand and use if you have access to a computer. It can be used for almost all kinds of statistical estimates.

Suppose you have a sample from an unknown distribution as well as an estimate of an unknown parameter. It is reasonable to assume that the empirical distribution of the sample resembles the unknown distribution. The empirical distribution is in fact the ML-estimate. If the random sample had been drawn from the empirical distribution, the standard error of the estimate would have been approximately the same. By drawing several samples from the empirical distribution and computing the standard error, you thereby get a good estimate of the standard error for the true distribution. This estimate is called the bootstrap.

There are several other techniques, where you choose a subsample from the observed variables and compare the corresponding estimates. Several methods are studied, including different versions of the jackknife, cross-validation, balanced half-sampling and random subsampling. These methods are compared with each other and with the bootstrap. The main concern is the bias and the variance of the error estimates, although confidence intervals and t-tests are mentioned.

In official statistics, the populations are often finite and the sampling plans complicated. Most resampling methods can be used in such situations after some more or less sophisticated modifications. Half-sampling was even developed for use in stratified sampling. The book, however, does not treat different sampling plans or experimental designs. This does not mean that the book is of no use for practicing survey statisticians. On the contrary, I think that the book might provide them with several good ideas.

The level of the book is such that an elementary knowledge of statistics is required but no deeper knowledge. Most people who are interested in statistics may benefit from reading the book – not only experimenters and students, but also practicing statisticians and theorists.

The book is short, but it covers the subject well. Such a book, of course, can be criticized for not covering everything. I would, for example, have liked to have seen chapters about the relation to permutation tests and about different sampling and experimental designs. In spite of this, I am impressed by the amount of important facts the author has managed to squeeze into the small amount of space.

Daniel Thorburn
University of Stockholm

Lindley, D. V. and Scott, W. F., New Cambridge Elementary Statistical Tables. Cambridge University Press, Cambridge, 1984, ISBN 0-521-26922-9, 80 pp., £2.

This set of tables replaces the well-known publication "Cambridge Elementary Statistical Tables" by D. V. Lindley and J. C. P. Miller, first published in 1953. The new set contains 28 tables compared with 10 in the old set.

According to the preface, "nearly all the tables have been newly computed for this publication and compared with existing compilations." Besides the usual tables of the normal, t, F, and χ^2 distributions, there are rather extensive tabulations of the binomial and Poisson distribution functions. Also included are a fair number of tables of the type required for nonparametric techniques. An unorthodox inclusion for the 1980's is the table of the Behrens-Fisher distribution.

A mild objection can be raised against the scantiness of the tables of random sampling numbers and random normal deviates. If thousands of readers use these figures over and over again, the randomness will be very poor!

Judging from the preface, the authors have suffered some pangs of conscience when including so many tables for significance tests in the new publication. They formulate their defense as "the student's need to follow prescribed syllabuses and to pass the associated examinations." Inevitably, misuse of a publication of tables may contribute to spreading the significance plague throughout the world, but we cannot blame the authors for that.

I recommend the entire set of tables, which are well chosen, designed with great care, and superbly printed.

Gunnar Blom
University of Lund
and
Lund Institute of Technology

Lewis, P. A. W., Endel, J. O., and Uribe, L., Introductory Simulation and Statistics Package. Wadsworth Advanced Books & Software, Belmont, California, 1984, ISBN 0-534-03741-0, viii+69 pp., \$39.95, 5 1/4" floppy disk included.

This package contains FORTRAN 77 subroutines to set up and analyze simulations on IBM PC using DOS 2.0 and Microsoft FORTRAN 77, Version 3.13, 8/5/83. Users of the package must call the subroutines from their own Fortran programs and a prerequisite to benefit from the package is knowledge of Fortran programming. Some knowledge of statistics is also necessary to understand and interpret the output from the programs such as boxplots, histograms and descriptive statistics.

The two-sided floppy disk includes nine subroutines, compiled under Microsoft \$NOFLOATCALLS compiler optimized with 8087 coprocessor support. All random number generators are fixed algorithms producing pseudorandom numbers. The algorithm used in this package,

$$U_{i+1} = (16807 * U_i)\mathrm{mod}(2^{31} - 1) ,$$

is described in Lewis, Goodman, and Miller (1969) and is used in several statistical mainframe packages including IMSL statistical library, SPSS, and IBM's APL. The algorithm produces over two billion pseudo-random numbers before repetition, and the numbers appear in such an order that they were "truly" random. This property is discussed and tested in Learmouth and Lewis (1973).

The generator can be used directly or indirectly by transformations to other distributions. There are four subroutines, for generating Normal, Gamma, Geometric, and Poisson random variables. The subroutine for generating normal random variables makes use of the Box-Muller (1958) sine-cosine transformation of Uniform (0,1) to Normal (0,1) random variables. If you need a Normal (μ, σ) variate, you must perform a user-defined transformation. This takes 1 minute, 36 seconds to generate 100 000 Uniform (0,1) and 4 minutes, 11 seconds for Normal (0,1) variates with support of the coprocessor. The Gamma generator produces Gamma(k,1) variables, where k can be chosen by the user. The algorithm for Gamma is an implementation of Schmeiser and Lal's (1980) acceptance-rejection scheme. The algorithm for generating Poisson random variates is described in Fishman (1978) and needs the Uniform (0,1) generator during linking.

The algorithm for Gamma (2,1) needs 6 minutes and 13 seconds for generating 100,000 numbers. The corresponding time for Poisson (5.0) is 4 minutes, 21 seconds with the 8087 coprocessor.

The simulation output can be analyzed by univariate and bivariate histograms providing graphical displays. The graphics for these programs requires only a standard dot-matrix printer. One routine performs a sort of data vector. According to the authors, the package could run without the 8087 coprocessor. However, the running time would then increase.

The manual for the package contains three parts. First, there is a general introduction to the programs; how the package operates and how to interpret its output. The second part describes the use of the program floppy disk, and the last part deals with the program parameters and their modifications.

A "typical" use of the package is the comparison of estimators with respect to bias, variance, and sampling distributions. If the estimators have known distributions, the package could be used in classrooms as a supplementary tool in courses in statistical simulation at the undergraduate or graduate level. As I mentioned earlier, the students must be able to write their own Fortran programs in an IBM PC-environment or compatible hardware.

The algorithms used in the package have been tested in other reports and the properties of the algorithms are well known. For that reason, I restricted the test of the output from the Uniform random generator. Five samples with a sample-size of 1000 were generated. The output was tested for fit to Uniform (0,1) by chi-square-tests. These tests do not reject the hypothesis of true Uniform distributions. The other subroutines have been run merely to see if they "worked," and they did.

There are several disturbing misprints and errors. For example: On page 6, there is a missing square root sign. On page 8, the denominator is erroneous.

References

Box, G. E. P. and Muller, M. E. (1958): A Note on the Generation of Random Normal Deviates. Annals of Mathematical Statistics, 29, pp. 610–611.

Fishman, G. S. (1978): Principles of Discrete Event Simulation. New York, Wiley.

Learmouth, G. P. and Lewis, P. A. W. (1973): Statistical Tests of Some Widely Used and Recently Proposed Uniform Random Number Generators. Naval Postgraduate School Technical Report, NPS55LW73111A.

Lewis, P. A., Goodman, A. S., and Miller, J. M. (1969): Pseudo-Random Number Generator for the System /360. IBM System Journal, 2, pp. 136–145.

Schmeiser, B. W. and Lal, R. (1980): Squeeze Methods for Generating Gamma Variates. Journal of the American Statistical Association, 75, pp. 679–682.

Sven Öhlén
Statistics Sweden

McLean, R. A. and Anderson, V. L., Applied Factorial and Fractional Designs. Statistics: Textbooks and Monographs Series, Vol. 55, Marcel Dekker, Inc., New York and Basel, 1984, ISBN 0-8247-7154-0, 373 pp., $78.

The subject of this book is factorial experimentation, a part of experimental planning.

The contents of the book are listed as follows:

As a special case, the book deals with fractional designs as seen from the title. This is a method of reducing the number of observations. As a consequence it is not possible to estimate as many effects as in a complete factorial experiment.

The book contains a number of schemes concerning different factorial and fractional designs. A rather odd feature, though, is that the appendices constitute over two-thirds of the volume. The appendices are lengthy and also contain many schemes with factorial and fractional designs but with much less text. Three of the four appendices are not written by McLean and/or Anderson.

As a matter of fact, rather few applied examples of factorial experiments appear in either the appendices or the rest of the book. The small number of applied examples seems to contradict the title of the book. For applied nonalgebraic examples of factorial experiments, see for example instead Snedecor and Cochran (1980) and its chapter on Factorial Arrangements of Treatments.

The book moves logically from the simple to the more complex. This reviewer thinks that an introduction to factorial experiments without fractional designs would have been better. As it is now, the book proceeds too fast to factorial and fractional designs. This, in combination with the earlier mentioned lack of applications, makes the book difficult to read.

The book is intended for practitioners in the field of factorial experiments and consequently the reader should be fairly familiar with factorial experiments.

The authors hope that the book will be used as a course book. It is suitable for such a purpose but for a knowledgable audience rather than those seeking an introduction to factorial experiments.

Reference

Snedecor, G. W. and Cochran, W. G. (1980): Statistical Methods. The Iowa State University Press.

Tomas Garås
Statistics Sweden

Parzen, E. (Ed.), Time Series Analysis of Irregularly Observed Data. Proceedings of a Symposium held at Texas A & M University, College Station, Texas, February 10–13, 1983. Lecture Notes in Statistics. Springer-Verlag, Berlin, 1983, ISBN 3-540-96040-6, 363 pp., DM 58.

This volume contains 16 papers which were presented at a symposium at College Station, Texas A & M University in 1983. The papers all consist of discussions of various problems related to time series analysis of irregularly observed data, that is, time series with missing data. The papers indicate the state of the art with respect to theoretical methodological research, as well as the relationship between methodologies and different fields of application. The present aspect of time series analysis is of great empirical importance, and thus of significant theoretical interest. The papers cover a wide variety of time series models, as well as methodological problems and applications.

Problems with regression analysis are discussed in papers by W. Dunsmuir, A. C. Harvey/C. R. McKenzie, R. H. Shumway and M. Hinich/W. Weber. Dunsmuir assumes, besides the presence of missing data, that the residuals are periodically correlated. Data on regressand and regressors are missing at the same points in time. The estimation techniques discussed are easily implemented. This simultaneous presence of two problems is very interesting both empirically and practically. Dunsmuir gives two applications, viz in the analyses of average monthly salinity and of average daily carbon monoxide. The Harvey/McKenzie paper involves discussions of the analysis of the empirically important dynamic linear regression model, written in state space form in the paper. They analyse various situations characterized by missing data, including the aggregation problem. Analysis by Kalman filtering and EM-algorithms are studied. Besides discussing estimation and smoothing in state space forms, Shumway analyses the estimation of frequency domain regression models. Of particular interest is his review of the literature in this problem area. The Hinich/Weber paper deals with the distributed lag models which are frequently applied empirically. An estimation technique based on the use of the Hilbert transform is introduced and numerically evaluated.

The time series models analysed in the volume are stationary in some papers and in others nonstationary. Furthermore, discrete as well as continuous time series models are considered. The paper by C. Ansley/R. Kohn includes state space representations of nonstationary ARIMA models. Nonstationary likelihoods are computed by modified Kalman filters. Problems related to the analysis of nonstationary discrete time series are also analyzed by G. Kitagawa. So called "penalty smoothing" is suggested as an interesting alternative for handling unequally spaced data. The methods are illustrated by a number of interesting examples from a wide variety of fields of applications, for example within the seasonal adjustment of economic time series.

A discrete stationary time series analysis is studied in another paper by Dunsmuir. He considers time series with missing data as being represented by "amplitude modulated" series. The time series studied are assumed to follow a finite parameter ARMA model. R. B. Miller/O. Ferreiro consider the problem of estimating missing values in discrete stationary AR(p) time series. Estimation of parameters in an AR(1) model when replacing missing values is examined by their conditional expectations. Multivariate first order autoregressive models are also studied by R. Jones. In particular, Kalman filtering is used to estimate transitions, etc. Illustrations taken from medical data are also given, as well as a list of computer programs.

Continuous time series are studied in a number of papers. D. Brillinger discusses estimation of spectral densities, for example. Special interest is given to the role of Fourier transforms for statistical inference. P. Robinson develops a general theory for analysis of continuous multiple, irregularly spaced time series. In particular, the series are given an infinite moving average representation. In V. Solo's paper, problems related to fitting continuous time series to discrete data are discussed. His paper includes an interesting review of the important problem of aliasing. An approach for avoiding that problem is also suggested. The paper by H. L. Weinert is to some extent related. Weinert introduces a Hamiltonian approach for continuous/discrete smoothing.

Nonparametric estimation of probability and spectral densities, covariance functions, etc., of stationary continuous time series (with irregularly spaced data) is thoroughly discussed in E. Masry's review. D. W. Marquardt/S. K. Acoff also consider a nonparametric technique for spectral analysis, viz the so called DSQE spectrum, originally suggested by Marquardt.

Several of the papers include applications and illustrations of certain techniques for time series analysis when data are irregularly observed, while concentrating on methodological issues. One paper, however, focuses on applications. A. D. Thrall/C. S. Burton present a detailed discussion of the problem of analysis in environmental science, where time series are often characterized by a substantial quantity of missing data.

The problem of irregularly observed data in time series analysis is of basic importance to applied researchers and to other consumers of techniques for time series analysis. For example, economic structural analysis, planning and forecasting are often heavily restricted due to the lack of knowledge on how to handle the missing data problems efficiently. That problem arises in temporal and contemporaneous aggregation. Thus, the Conference on Time Series Analysis of Irregularly Observed Data and the present proceedings volume is of great importance and interest, not only to theorists but also to practitioners. The papers, all of which are refereed, provide an excellent review of the state of the art of time series analysis of irregularly observed data.

Anders H. Westlund
University of Umeå

Printed in October 1985.