

Statistical Data Analysis Handbook. By F. J. Wall. ISBN 0 07 067931 2. McGraw-Hill, New York, 1986. 546 pp. \$49.50.

The author's aim in writing the book is to show the naive data analyst how to use statistical methods by focusing his attention and experience on a wide variety of practical problems. However, do not be misled by the title of this book. The book is a collection of real data and some statistical treatments on them with varying degree of detail. The book is not sufficiently

comprehensive to be used as a reference book for the practising statistician. As it is, the book does not serve as a handbook in data analysis.

The book contains five parts. The first three parts present the topics in the context of analysing detailed practical problems.

Part I—'Simple comparisons'—deals with the t test for independent and paired data, and ANOVA. For these parametric techniques the related nonparametric procedures are also discussed. Analysis sequence flowcharts are also given. Part II—'Linear relationships'—gives an introduction to simple, polynomial and multivariate linear regression, and in Part III—'Comparison of relationships'—two or more than two linear relationships are discussed. Part IV—'Supporting topics'—is a survey, reviewing introductory statistics for those who have not already had exposure to statistical theory. Part V—'Tables'—contains 20 tables mostly arranged in terms of the upper tail probabilities (P values).

The first four parts are divided into 20 chapters paged within each chapter. Some chapters are very brief. I would have preferred Chapter 14, which covers transformations, to be omitted since it is too brief. Thus all the references to this chapter (e.g. pp. 2.13–2.18) throughout the book should be omitted. Certain important topics have not been adequately covered or discussed. Particular omissions which might be seen as important include: analysis of residuals, remedies for violations of regression assumptions, the treatment of outliers, autoregressive error structures for time series data (example 10.1) and the Durbin–Watson test, and the effect of sample size on the meaning of the P value which is an issue (Lindley and Scott, 1984) since some of the data sets are small. These topics are closely intertwined, however, and the model building chapters are incomplete without the residual analysis.

In addition, there are a few mistakes or ambiguities. These could cause real problems for a student approaching the subject for the first time. These topics include the following.

- (a) The computation of Lilliefors statistic for data sets having some observations with equal values (e.g. Table 15.6) is incorrect. This error is caused by miscalculating the sample probabilities. Although this correction has no impact on the results, I found it rather tedious to compute this statistic and prefer to use the readily available acceptance graphs, e.g. Iman (1982), for this test.
- (b) The test for equality of variances assumes normality which must be tested first. All the given flowcharts must be corrected to reflect this fact.
- (c) Some of the given nonparametric procedures are for the case when there are no ties in the data and thus could be misleading. Clearly, a unified approach, e.g. Meddis (1984), could condense the presentation of nonparametric procedures.
- (d) The chi-square test for contingency tables is performed on data sets which do not have sufficient expected counts. This contradicts the advice elsewhere in the book.

These problems could have unfortunate consequences if the handbook were being used as a cook book. These should be corrected in any later edition of the book. Some unimportant typographical errors slipped through the net.

The book has many virtues to combat the disappointments which this review has highlighted. The author avoids simple, artificial, dreamt-up kinds of example in favour of real data sets. This is to be applauded, although some of them are classical and treated better by others. The book is well stocked with detailed descriptions of how to carry out the analysis of a data set. This allows users to repeat the calculations on their own data. The hallmark of the book is that there is a definite emphasis on conditions to perform statistical testing and the usage of the P values. Without being too enthusiastic I am happy with this book. It is easy reading and provides interesting introductory material which is suitable for students with a wide range of backgrounds particularly with students who have limited mathematical backgrounds. The book is aimed at the statistically naive market and its intended purchasers are practitioners who have done at least a first course in statistics.

References

- Iman, R. (1982) Graphs for use with the Lilliefors test for normal and exponential distributions. *Ann. Statist.*, **36**, 109-112.
- Lindley, D. V. and Scott, W. F. (1984) *New Cambridge Statistical Tables*, p. 3. Cambridge: Cambridge University Press.
- Meddis, R. (1984) *Statistics using Ranks, a Unified Approach*. Oxford: Blackwell.

H. Arsham
University of Baltimore