# Review of Neil Salkind's 4th edition of "Statistics for People Who Think They Hate Statistics" (2011, Sage)

Gabriel Donleavy

University of Western Sydney, Australia

The book reviewed is an introductory text on descriptive and inferential statistics for students in the social sciences who need to do a research project or who need a basic level of statistical understanding.
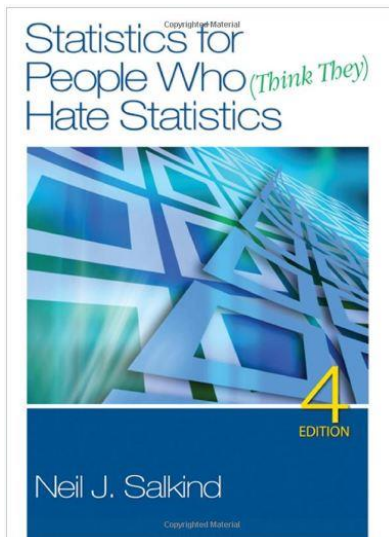
The book's introduction defines statistics as "the science of organizing and analysing information to make the information more easily understood". The goal of the book is "to provide you with an understanding of these basic tools and how they are used". The book thus does not aim merely to introduce statistics, but rather to provide readers with an understanding of basic statistical concepts, an understanding which involves both their calculation and their usage. This calls for a higher standard of critical analysis than what would be appropriate for a review of a basic introduction to the field, for "understanding" is not a modest goal. The book "makes no presumptions about what you should know before you start and proceeds in slow, small steps letting you pace yourself." (p.11). Indeed, the author tries really hard to put students at ease. There are many examples fully worked.

In this review, I will focus on the extent to which the book achieves its goal of having the novice understand the statistical concepts covered. I will consider whether the scope of the book is appropriate to equip its intended audience with the tools to process research data and whether the book explains key statistical concepts better or worse than its major competitors do. Finally, I will consider whether I could recommend this book at all, in its entirety or for some things but not for others.

Aims

There are many introductory statistics books on the market. Any entrant into this market needs to fill a particular niche. This one hopes to fill the niche implied by its name – the niche where people have to take statistics because they are obliged to do so but who do not like the prospect, whether because of innumeracy or other reasons. Because it is aimed at unwilling novices, the book not only has to unpack thoroughly all technical terms and formulae but it also has to explain them in a language plain enough to stop a beginner from becoming frightened, confused

or bored. The writer must have succeeded to some extent in occupying his intended niche; else he would not have achieved four editions so far.

Emphasis and Scope Issues

The main criticism of the book's emphasis is that it has far too much on parametric statistics and far too little on non-parametric statistics. Chi square is dealt with adequately, but the processing of ordinal data and of non-normal distributions is disappointingly thin and could not be used by a student who needs techniques from those areas to process data for their project. These areas are done much better in the classic books by Blalock (1979) and Spiegel (2011), and many students will be dealing with skewed and lumpy data, so they actually need nonparametric more than they need parametric. However, Salkind is not alone in this emphasis on the parametric, unfortunately.

Salkind's concern for strict exactitude in definition is sometimes at the expense of ease of understanding by innumerate students. For instance, "A small "n" represents the sample size for which the mean is being computed. A large "N" would represent the population size. In some books and in some journal articles, no distinction is made between the two." (p. 21). Why does a student at this stage really need that last sentence? Similarly, in contrasting sample statistics' use of roman letters with population parameters' use of Greek assumes readers already know both alphabets (p. 24).

Part 2 on descriptive statistics is good on means and medians but inadequate on skew and kurtosis whose presence can often render parametric techniques quite inappropriate. Other issues of concern include:

- Page 92's Table 5.3: the header "correlation coefficient shopping anyone?" lists five types of correlation coefficients (phi, rank bi-serial, point bi-serial, Spearman and Pearson) and shows the measurement scales applicable to each of the two variables for which each correlation coefficient is most appropriate. This is much too important just to be dropped in without development.
- Appendix E on basic maths: it is good but ought to be a preface rather than an epilogue to the book.
- Statistical software: SPSS is used throughout the book, but the book had earlier raised the expectation of also using Excel and we never hear again about that. Chapter 19's trip through statistical software is clear but why does it omit SAS and STATA when they are so common in social science and business research?
- Chapter 17: this is the one and only chapter that deals with non-parametric tests but does not do so clearly or adequately. It is given four smileys which is too generous, considering that it is not clear which attributes of normality are essential for parametric tests and which are not.

Finally, quite a serious omission in Chapter 14 is the failure to explain the general linear model which is nevertheless heavily introduced (p. 249); similarly the selection of more advanced inferential in chapter 18 has no obvious rationale. It mentions MANOVA, ANCOVA and factor

analysis but makes no mention of PROBIT, LOGIT, discriminant analysis, cluster analysis nor multidimensional scaling, though these are all important in business and social research.

Clarity of explanations

Many paragraphs are accompanied by icons in the margin representing the paragraph's content type (introduced at p. 14). This is a good idea which the author uses well throughout the book. Then we have the smileys representing a difficulty index from 1 to 5. I would have had a 3 point scale and used frowns for hard parts and zero for neutral. This would have made the scale easier to understand and would avoid the spurious precision of a 5 point scale for something as subjective as ease and difficulty. For it is quite hard to gauge relative difficulty for innumerate students and anything rated as hard or very hard by the author may increase rather than decrease a new student's fear of the relevant topic.

Commendable explanations and treatments are many and are, in every case, at least as good as, and sometimes better than, the explanations in most other textbooks. Salkind excels in the following areas: Chapter 4's treatment of pictures, diagrams and charts, mean deviation (p. 40), measurement scales (pp. 103-105), sampling errors (p. 128), null hypotheses (p. 129), chance as explanation of results (p. 130) and research hypotheses (p. 131), the asymptote of normal curves (p. 143), the transition from z scores to probability (pp. 152-153), why type 1 errors affect test design more than type 2 (p. 170), the implication of obtained values being more than critical ones (p. 176), the f test (p. 222), factorial ANOVA design (p. 224), ANOVA's superiority over multiple t tests (p. 232), main effects and interactive effects for 2 way ANOVA (p. 240), how ANOVA spots effects that t tests miss (p. 245), significance versus meaningfulness (p. 260, in contrast to the confusing introduction of the topic on pp. 170-171), regression's line of best fit (p. 271), standard error of the estimate (p. 275), multiple regression and its selection of independent variables (pp. 279-280), chapter 21's ten commandments of data collection, appendix A's guide to SPSS.

Salkind is not quite so clear, however, on several topics, starting with skew which is dealt with much too briefly although clearly enough (p. 27). The treatment of skew is quite a good test of the quality of an introductory statistics book, because it is the moment primarily responsible for partitioning datasets between the normal and the non-parametric and it is quite easy to explain, but rarely done well in the books. Second, when standard deviation is introduced, the book fails to explain why the formula has the content it has (p. 40). There is a very good clear explanation of mean deviation but no mention of why it is hardly ever used (p. 42). It is also hard to understand why the author says (p. 45) that variance will be dealt with later in the book but then it is not; and this omission may magnify its difficulty for some readers.

Further, all of chapter 9 on significance testing is significantly less clearly dealt with than other techniques and needs simplifying for the fifth edition; how to do one tailed tests is not fully enough explained (p. 213, although the much later treatment on p. 258 differentiating one from two tailed tests is in fairness quite clear); degrees of freedom are not well explained in chapter 15. Penultimately, one smiley for chapter 16 on regression is merited, but, for the previous two chapters on ANOVA, it is self-deprecating in a way that may unnecessarily scare the novice, as both chapters are clearer than most of the rest of the book.

Finally, not very user friendly is appendix D which has answers to practice questions. The answers are fine, but the impression is spoilt by the occasional interpolation of the phrase "on your own" in quite a few of the answers without saying it first in the questions within the chapters. The point is that the better the model answers are (and Salkind's are very good indeed), the more frustrating it is to be told some are not given and we have no reason why not.

Comparison with competitors

Gonick and Woollcott (1993) "Cartoon Guide to Statistics" might be the simplest introduction for some students who learn more easily from the visual than from the verbal, but it has not had a new edition for quite some time and has had mixed reviews. It is good on descriptives and display, probability, and Bernoulli trails, but the reader still needs to work, for cartoons do not offer osmosis learning. Salkind compares well with this book.

Spiegel (2011) is well reviewed on Amazon.com; it is long established and very clear text. It is better than Salkind on skew, Poisson and binary distributions, sampling theory and applications and time series analysis, but is not better than Salkind on non-parametrics and more advanced areas like the general linear model. Some reviewers thought some numeracy was a pre-requisite for this book. Weiss's (2010) Introductory Statistics is a full textbook, with many good Amazon.com reviews, and with the same coverage and easier language than Salkind, but it too is weak on non-parametrics and inferential analysis beyond ANOVA.

Wasserman (2003)'s "All of Statistics" is introductory but comprehensive on both inferential and on descriptive, and its language is accessible if not crystal clear. I think it is rather good, but Salkind is easier for the beginner.

Griffiths (2009) "Head First Statistics" looks more novice friendly than the others from a distance, as it is full of puzzles, quizzes, stories and real world examples. It does all the main distributions, probability, and inferentials up to regression and chi square. Some reviewers found it silly, others clear and engaging; it has some spelling errors and wrong answers. It is very good on the meaning and appropriate use of each statistic, but a shade too careless relative to the other books, including relative to Salkind.

Blalock (1979) is a classic text, still easily bought; it is appropriate for the serious research student and has no carelessness in it, but it is, regrettably, too advanced for the novice who is compelled to take a statistics unit in their course, whereas Salkind is not at all advanced and is successful in its targeting the beginner.

Conclusion

Salkind's book is well above average for an introductory statistics textbook and may be the best for its integration of SPSS and its explanation of t tests and ANOVA. It could usefully be followed by a book that builds on what he has already in print and which takes the student from beginner to research capable status. The fifth and subsequent editions will hopefully address the omissions mentioned above. All in all, this book can be recommended as helpful and clear for the beginner.

References

Blalock, H. (1979). *Social statistics*. McGraw-Hill.

Gonick, L., & Smith, W. (1993). C*artoon guide to statistics*. HarperCollins.

Collins. Griffiths, D. (2009). *Head first statistics*. O'Reilly Media.

Rumsey-Johnson, D. (2010). *Statistics essentials for dummies*. Wiley.

Spiegel, M., & Stephens, L. (2011). *Schaum's outlines: Statistics*. McGraw Hill.

Weiss, N. (2010). *Introductory statistics*, 9th Ed. Pearson Paperback.

Wasserman, L. (2003). All of statistics. In T*exts in Statistics*. Springer.

Citation

Disclaimer

Opinions expressed in this article are those of the author, and as such do not necessarily represent the position(s) of other professionals or any institutions.

Bio

Gabriel Donleavy is a Professor of Accounting at the University of Western Sydney. He has published six books, 37 refereed scholastic articles in academic journals (including the Journal of Business Ethics, the British Accounting Review, and the International Journal of Accounting and Economics), and 70 papers and monographs in the fields of business education, accounting and business ethics. Professor Donleavy has worked in universities in England, Singapore, Hong Kong, Macau and Australia and has designed, reviewed, validated, and accredited business and accounting courses with the national accounting bodies in England and Hong Kong. His academic career includes three deanships and two head of school positions. Professor Donleavy was an organizer of Hong Kong Amnesty International before and after Tiananmen Square protests in the late eighties and a National Trustee of England's Citizens' Advice Bureaux at the turn of the century. At present he is a member of the New South Wales advisory board of the National Seniors' Association. Professor Donleavy may be reached at g.don@uws.edu.au