Book review

Statistics and Experimental Design for Psychologists: A model comparison approach

Peter M. C. Harrison

This article reviews a new statistics textbook by Rory Allen. The book teaches the fundamentals of statistics and experimental design at an undergraduate psychology level. The material is taught from a model comparison perspective, which Allen argues to be more intuitive than traditional null hypothesis significance testing (NHST) approaches. A key feature of the book is an elegant and novel geometric formulation of Fisher's F, the test statistic in analysis of variance (ANOVA). Allen's contribution should provide a useful reference for psychology students, especially if future editions incorporate a wider range of statistical methods.

STATISTICS TEXTBOOKS are traditionally built around null hypothesis significance testing (NHST). NHST begins with a *research hypothesis* that certain structure exists in a data-generating process. This hypothesis is inverted to form a *null hypothesis*, stating that no such structure exists. Observed datasets are analysed in terms of a *test statistic* (e.g. the *t*-statistic in a *t*-test), which is a function of the dataset with a known distribution under the null hypothesis and a different distribution under the research hypothesis. An unlikely test statistic under the null hypothesis provides support for the research hypothesis. The degree of support is quantified by the *p*-value, defined as the probability of observing such an extreme test statistic under the null hypothesis.

There is something counter-intuitive about validating a research hypothesis by disproving a null hypothesis. In this new textbook, Allen argues that this aspect of NHST is responsible for much confusion in psychology students. His proposed solution is for statistics educators to de-emphasise NHST in favour of a model comparison approach. Here both null and research hypotheses are represented by statistical models, and the hypotheses are compared by testing the relative fit and complexity of their respective models.

Statistical models can be compared in many ways. A particularly attractive approach is Bayesian inference, which brings many potential advantages over orthodox statistics (e.g. Wagenmakers, Morey & Lee, 2016). The statistics educator is faced with a dilemma, however: even if they dream of a Bayesian future for psychology, the vast majority of past and present psychological literature is still rooted in NHST. On this basis, Allen seeks a compromise; he retains the orthodox statistical methods of NHST, but reinterprets these methods from a model comparison perspective.

With model comparison as its unifying perspective, this 448-page book covers fundamentals of statistics and experimental design as they might be taught in introductory courses for undergraduate- or master's-level psychologists. Taught statistical techniques include descriptive statistics, normal distributions, independent-samples *t*-test, one-way analysis of variance (ANOVA), multifactorial ANOVA, and repeated-measures ANOVA. Each statistical technique is situated within the wider domain of experimental design: for example, the independent-samples *t*-test is motivated as an analysis method for the randomised controlled trial. Theoretical descriptions of statistical techniques are followed by detailed tutorials for applying these methods using the popular statistical software package SPSS. These matters of statistics and experimental design are complemented by detailed attention to the philosophy of the scientific method and its historical development, including Occam's razor, Bacon's inductive method, Popper's hypothetico-deductive method, and Fisher's NHST.

The inferential statistics taught in this book are all special cases of ANOVA. Allen provides an elegant geometric interpretation of the ANOVA test statistic, Fisher's *F*, that is particularly suited to the model comparison approach. Given the novelty of this approach, it seems worthwhile to summarise it here.

The essential elements of the approach are best illustrated by its application to one-way ANOVA. The aim of one-way ANOVA is to test for grouping structure in a set of observed data. Under Allen's formulation, this is approached by comparing three competing models. The first is the *null model*, which models the observed data as one normally distributed group. The second is the *saturated model*, which models each observation as a separate group. The third is the *candidate model*, a compromise between the null and saturated models that models the observed data as a collection of normally distributed groups, with these groups defined with reference to the research hypothesis.

These three models are plotted on a Cartesian plane with number of model parameters on the horizontal axis and lack of fit sum of squares (LOFSOS) on the vertical axis (Figure 1). Number of model parameters is a measure of model complexity, whereas LOFSOS is a measure of model fit, with lower LOFSOS indicating better fit.

Model comparison involves a compromise between model fit (LOFSOS) and model complexity (number of parameters). The null and saturated models illustrate opposite extremes on this spectrum: the null model achieves minimal complexity at the expense of minimal fit, whereas the saturated model achieves maximal fit at the expense of maximal complexity. Allen connects these two points

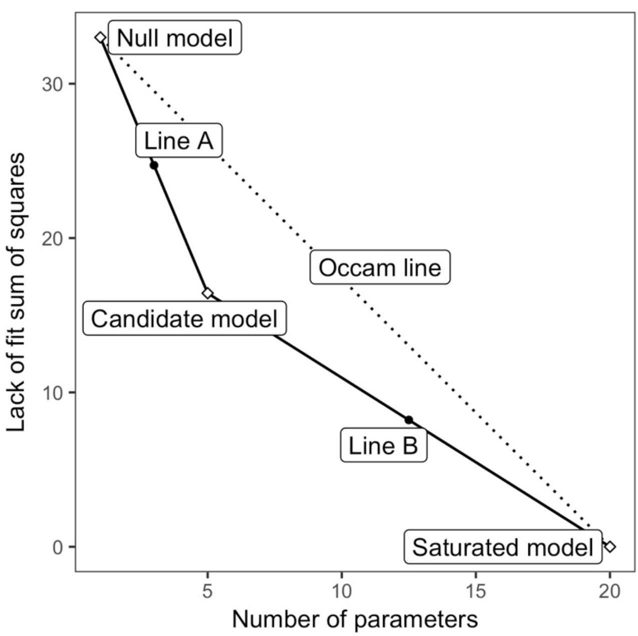

Figure 1: Geometric interpretation of the ANOVA test statistic, Fisher's F, after Allen (2017).

with a line which he calls the *Occam line*; this line is argued to represent all models equally uninformative as the null and saturated models. Intuitively, therefore, models that fall significantly below the Occam line should be preferred to the null and saturated models.

As it turns out, this geometric condition can be linked to the traditional ANOVA F-test. The F statistic can be calculated from Figure 1 by taking the gradient of Line A and dividing it by the gradient of Line B. As a result, a necessary and sufficient condition for falling under the Occam line is an F statistic greater than one. Traditional ANOVA NHST selects the research hypothesis if the F statistic is sufficiently large to be implausible under the null hypothesis. Such F values are always greater than one, by nature of the F distribution. Statistically significant F tests therefore always correspond to models below the Occam line.

This geometric formulation of ANOVA plays a central role in the book. It is first presented in a general chapter on model comparison ('Comparing different models of a set of data', pp.31–61), and is revisited each time a major new technique for statistical inference is presented (independent-samples t-test, p.223; one-way ANOVA, p.266; multifactorial ANOVA, p.332; repeated-measures ANOVA, p.376). This repetition usefully emphasises the generality of the approach, and helps to unite the different techniques under the same model comparison umbrella.

The main pedagogical advantage of the diagram is the intuitive illustration that model construction must balance model fit and complexity. This principle is less apparent in the traditional variance partitioning approach to ANOVA.

While the geometric illustration of Fisher's F is (as far as I know) unique, the conceptual explanation is similar to some other textbooks. For example, both Judd, McClelland and Ryan (2017) and Maxwell, Delaney and Kelley (2018) construct the F statistic in terms of the proportional change in squared error normalised by the change in the number of model parameters.

These recommended textbooks (Judd et al., 2017; Maxwell et al., 2018) in fact share a deeper similarity with Allen's contribution: both present orthodox statistics from a model comparison perspective. Both are excellent books and important competitors with Allen's work, and indeed Allen recommends both in the 'Further Reading' section. Judd et al. (2017) is particularly relevant, as it is targeted towards a similar readership as Allen's book. Maxwell et al. (2018) is aimed somewhat higher, and assumes that the reader has taken at least one statistics course.

The clearest advantage of these competitors is the broader range of statistical techniques that they cover. Allen's book provides a rather limited set of techniques, essentially limited to simple ANOVA variants, that would be insufficient for many master's-level statistics courses. Particularly striking is the lack of techniques for dealing with continuous predictors, such as correlation, regression, and ANCOVA. Allen explains the limited scope of the book in terms of the difficulty of simultaneously addressing both statistics and experimental design, writing that 'a book dealing fully with both areas would be far too long for a single volume' (p.viii). I think Allen is being overly pessimistic – the runaway success of Field's 'Discovering Statistics' series (e.g. Field, 2017) has demonstrated that students can stomach long textbooks as long as they are tempered by sufficiently many personal anecdotes.

Nonetheless, Allen's book has several unique selling points. Unlike the two model comparison textbooks mentioned earlier (Judd et al., 2017; Maxwell et al., 2018), it is targeted specifically at psychologists. This allows Allen to focus on research designs that are particularly common in psychology (e.g. t-test, ANOVA), and to give worked examples that are particularly relevant to psychologists. Acknowledging the fact that many psychology students do not come from strong mathematics backgrounds, the book adopts an approachable style with relatively few equations, placing instead a greater emphasis on the philosophy of the scientific method.

The geometric interpretation of the ANOVA is a good example: it helps the student to appreciate that model comparison involves a balance of fit and complexity, without the need to memorise equations. The approachability of the book is further aided by the way that Allen motivates each statistical technique by realistic research situations. Detailed SPSS tutorials are provided to help the reader put these statistical techniques into practice. Allen also has the interesting idea of providing Excel tutorials that reconstruct the calculations involved in the various statistical tests. This gives students an important low-level perspective of the techniques without requiring laborious hand calculations or knowledge of a fully-fledged programming language.

In the book's preface, Allen writes 'the website associated with this book contains extensive further material, not only extending the coverage in the earlier chapters, but also dealing with multiple regression, ANCOVA and exploratory factor analysis' (p.viii). From this I expected further chapters on these different techniques, presumably presented from the book's model comparison perspective. Instead I found a repository of narrated PowerPoint presentations which do not align closely with the book's chapters and do not take the book's model comparison approach. Allen writes that he is currently converting these materials to the model comparison approach, but cannot yet promise a completion date. Hopefully these updates will be completed soon.

Alongside example datasets, the companion website also includes supplementary materials for each chapter. These materials contain some useful gems, including some colourful questions and worked answers. However, not all of these supplementary materials are signposted in the book's main text, meaning that readers might pass them by. The delivery of these materials is not ideal, either: the reader must download a compressed archive for the chapter of interest, and inspect it on their local computer for relevant material, typically in the form of Word documents. Browsing these materials would be much easier if they were instead hosted as web pages connected by hyperlinks.

The limitations of the companion website should be easy to address in future versions of the book. The book's scope – from descriptive statistics to repeated-measures ANOVA – could stay the same, and the book would be a solid contribution to psychology education. In this context I can see it functioning well as a course textbook for undergraduate psychologists, or as preliminary reading for a postgraduate psychology course. However, I think that the book would come into its own if it expanded its statistical toolbox to match that of other standard textbooks such as Field (2017) and Judd et al. (2017). The book could then be used as primary material for postgraduate research methods courses, or indeed as a general reference book for early career researchers.

Correspondence

Peter Harrison
PhD student, School of Electronic Engineering and Computer Science,
Queen Mary University of London.
Email: p.m.c.harrison@qmul.ac.uk

Acknowledgements

The author is supported by the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1).

References

Allen, R. (2017). *Statistics and experimental design for psychologists: A model comparison approach.* London: World Scientific.

Field, A. (2017). *Discovering statistics using IBM SPSS Statistics.* London, UK: Sage.

Judd, C.M., McClelland, G.H. & Ryan, C.S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond.* New York, NY: Routledge.

Maxwell, S.E., Delaney, H.D. & Kelley, K. (2018). *Designing experiments and analyzing data: A model comparison perspective.* New York: Routledge.

Wagenmakers, E.J., Morey, R.D. & Lee, M.D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science, 25*(3), 169–176. https://doi.org/10.1177/0963721416643289