Review of *Modern Experimental Design* by Thomas P. Ryan

Though laymen may scoff at the notion, we professionals know that statistics texts can be very good reads. Careful elucidation of complex principles, well-chosen examples revealing general behaviors and the occasional idiosyncracy, topics organized to enhance progressive understanding in accord with a well-constructed foundation, precise use of terminology; these are the hallmarks of texts one enjoys reading, learning from, and teaching. These are also among the many areas in which *Modern Experimental Design* fails.

As you have no doubt surmised, this is going to be a negative review. This is especially unfortunate given the surfeit of useful material promised by a scan of the chapter and section titles. A detailed critique will follow, but let's begin by noting that these pages are difficult to turn. Not physically, of course; the problem is with a sometimes stream-of-consciousness style that explains too little while saying too much, so that plowing through yet another page demands setting aside a rising mountain of frustration. Some of this is due to the low technical level, but much is simply for want of structure, as in not identifying and separating key from side issues.

The Preface (page xv) claims that this book "would be suitable for an undergraduate one-semester course in design of experiments" and that there is "enough advanced material for the book to be useful as a reference book in a graduate course taught to statistics majors." The back cover claims that it is "A complete and well-balanced introduction to modern experimental design" that is a "well-rounded learning tool for beginners as well as a valuable resource for practitioners." These remarks bear insufficient semblance to the text, which quite different from being a general introductory treatment of design, is almost entirely industrial in its outlook. Especially, the poor treatment given the basics of the subject, and the scant attention paid the problems and requirements of non-industrial applications, make it ill-suited for beginners. The lack of depth afforded the "advanced material" likewise makes it unsuitable as a direct reference for graduate students, though it does cite many papers that would be useful.

Who, then, comprises an appropriate audience for this book? These individuals should have some experience with statistical methods combined with a strong interest in, and more than a passing acquaintance with, industrial applications requiring manipulation of numerous factors. On the whole these will be working engineers with sufficient experience to recognize the serious design problems they face. A good number of topics have been brought together at a suitably non-technical level so to make an accessible reference book for this group. One could easily see two or three day-long short courses for engineers based on this book. But let me repeat, it is wholly unsuitable for statistics majors at either the undergraduate or graduate level.

The broad structure of this book is not unlike many other design texts. The first few chapters cover basic concepts, the completely randomized design, blocking, and full factorials. Succeeding chapters branch off in many directions as detailed below.

The difficulties begin immediately in Chapter 1. We are told (page 6) that "Randomization is, loosely speaking, the random assignment of factor levels to experimental units. Ideally, the randomization method described by Atkinson and Bailey (2001) should be used whenever possible, although it is doubtful that hardly any experimenters actually use it." That last statement will shock a plethora of statisticians and other experimenters who have performed many randomizations exactly as described on page 57 of Atkinson and Bailey (2001), as taught in good design courses everywhere, and as easily implemented in virtually any statistical software (eg. PROC PLAN in SAS). While randomization is presented (page 6) to "reduce the possibility of confounding effects that could render an experiment practically useless," none of the other important reasons for randomization is mentioned. Only through these omissions and a narrowly industrial focus can one, as done here (page 7), invoke Box's (1990) otherwise myopic advice that randomization "should not be used if it more than slightly complicates the experiment, but there is a strong belief that process stability has been achieved ...." What follows is a murky discourse on how not randomizing can lead to unreliable results in the presence of time-related nonstationarity for sequential experimentation; this sudden introduction of undefined concepts will surely confuse anyone new to the subject. These first few pages typify the relentlessly industrial focus as well as a penchant for discussions that require experience far beyond that of an undergraduate statistics student.

On page 13 readers are again told "true randomization will often not be possible," apparently meaning that *complete* randomization may not be possible. While useful, statistically coherent (i.e. "true") randomizations for a wide variety of designs are well-known (consider the text by Hinkelmann and Kempthorne, 1994, or the more technical paper by Bailey and Rowley, 1987), you will not discover that from this book. See also the discussion (page 148) of restricted randomization in the context of hard-to-change factors, which does not even consider the types of restrictions handled in some of the cited but tangentially related papers. Certainly no statistician conversant in design theory would say that "restricted randomization and the problems caused by it might seem to have been seriously considered only since the late 1990s...." Restricted randomization *solves* problems, but like any statistical technique must be properly used. It can only be said to *create* problems when the relationship between randomization, model, and analysis is ignored. If that is the author's point, fine, but this book falls well short in making these connections.

The three pages devoted to choice of factor levels also contain missteps. Consider a factor X with two levels and a response Y whose variance is the same at every X. For a linear relationship, the text tells us that $\mathrm{E}(\bar{Y}_1 - \bar{Y}_2)$ increases with $X_1 - X_2$, but (page 17) "that spreading out the X-values will increase the spread of the Y-values, which will increase the denominator of the test statistic." This is odd, since variance homogeneity ensures the distribution of the error mean square is not a function of the factor levels. Fixed and random effects for factor levels are briefly mentioned here; the latter receive spotty, inadequate attention throughout.

The introduction (page 8) to the idea of replication is anything but transparent. Unless readers are already aware of the issue being discussed, they will not gain the important distinction between replication and multiple readings attempted here, nor understand the dangers. Hopefully no one will learn to say that an experiment was not properly replicated because "the experimental material was not quite identical between replicates, or even within replicates." A fundamental point, seemingly lost here, is that variation in experimental material is a central concern of design techniques, beginning with the completely randomized design.

Chapter 2 is devoted to the completely randomized design. Both fixed and random treatment effects are considered, though the only analysis offered for random effects is the basic ANOVA test. Sum-to-zero constraints are imposed on the model (rather than on the normal equations alone). Analysis of Means (ANOM) is presented as an alternative to ANOVA for the fixed effects setup. The cursory treatment of multiple comparisons tries to communicate the basic ideas. Some attention is payed to residuals analysis, but the author plainly states (page xvi) that "Although this book has more analysis than most books on design of experiments, it is not intended to be a handbook on data analysis." The comparison to other books is debatable, but the intention holds firm throughout.

Chapter 3, which takes up the topic of blocking, is unsatisfactory. Far too much effort is expended on the supposed defects of the RCBD, with very little being said of their strong advantages. The relevance of block-treatment interaction to analysis depends on how blocks are modeled, a point mangled if not missed altogether (page 58). The efficiency of the RCBD to the CRD is presented as an unreliable measure (page 61), there otherwise being no discussion of blocking's ability to reduce experimental variation. Among the unusual conclusions are (page 59) that "it should be apparent that there will erroneously be a failure to detect a significant difference in the means of a factor in many if not all applications of randomized block designs because of the number of blocks that are typically used." Also, "the usual unreplicated randomized block design should generally be eschewed in favor of a replicated design." The latter refers to what is commonly called the generalized randomized complete block design, or GRCBD. Though alluded to here, the GRCBD is never named nor explicitly handled. In ANOVA tables for block designs, the F-test for block effects is employed without mentioning the surrounding controversy (this test is not valid in the model generated by randomization).

The section on incomplete block designs belies the book's title, as it would be considered modern no later than about 1975. The focus is on balanced incomplete block designs and partially balanced incomplete block designs. Passing mention is given to other design classes such as lattices and α-designs, without mentioning the key property (resolvability) for which these designs are typically selected.

If RCBDs are not to be recommended, then surely the Latin squares must fare even worse. Fortunately we are given (page 77) the fairer assessment that "despite the somewhat gloomy picture painted at the end of section 3.3.1 regarding the likelihood of Latin square assumptions being met, the design has been used successfully in various applications." Make what you will of the statement (page 72) that "Since the assumption of no interaction probably won't be met exactly, it seems highly questionable to say that Latin squares are equivalent if the rows or columns of one design can be permuted so as to create the other design." Graeco-Latin squares and Youden squares are also introduced in this chapter, but again nothing modern for two or more blocking factors.

Chapter 4 explores the ideas of full factorials, both replicated and unreplicated, through the 2^k designs. A prominent feature is the considerable attention given to examining simple effects (here called *conditional effects*) for understanding factors in the presence of interaction. Effect plots are used for unreplicated experiments, though little guidance is offered on their interpretation. Pseudo-standard errors are discussed and the importance of effect sparsity emphasized. A one-page section is devoted to models with all effects random, where the poor power of F-tests is noted. This is done only for the 2^k, missing the important power questions for inference on variance components and their dependence on numbers of levels selected.

The muddled handling of randomization and replication continues. The text offers no probabilistic framework in which to evaluate this statement (page 149): "Using an experimental run sequence determined from cost considerations is of course a form of restricted randomization and thus makes the use of hypothesis tests using ANOVA rather shaky, although they still might be used, recognizing their limitations." We are again reminded (page 123) that "true replicated experiments rarely exist in practice...."

Chapter 5 begins the study of fractional designs, a major topic of the book, here for 2-level factors. A fractional factorial design is implicitly defined as what in modern terminology is called a regular fraction (any two effects are either fully aliased or orthogonal), and thus are considered to be a subset of the orthogonal arrays (page 229): "... all fractional factorials are orthogonal arrays but not all orthogonal arrays are fractional factorials." As with much of the technical material from here on out, defining relations are introduced in an informal manner, chiefly through example. A section is devoted to minimum aberration and number of clear effects in regular fractions, including an unenlightening discussion of small cases. Two case studies are presented for blocking of fractions, but no one will learn how to implement this technique from this presentation. The section on projective properties goes no further than mentioning that resolution R implies a full factorial in any $R-1$ factors, ignoring an abundance of important, modern results such as those of Cheng (1995). Other topics include foldovers, $\frac{3}{4}$-fractions, and fractions for natural factor subsets. There is a useful section on software capabilities relative to the techniques of this chapter. Here and elsewhere software comments are also mixed with substantive topics, usually to less beneficial effect. Telling us (page 192) that fractions produced by MINITAB "are probably minimum aberration" serves no apparent purpose - facts are needed for software, not conjecture. Software is strongly emphasized throughout the book; the primacy of software is reflected in comments such as (page 146) "The user of software for experimental design does not have to be concerned with how expected mean squares are obtained ..." and (page 188) "... new methods such as this one are not likely to be used to any extent unless they are implemented in statistical software."

Fractional designs for factors with more than two levels are the topic of Chapter 6. Classical design construction methods here are presented by example; design tables are referenced. Did you know (page 264) that "probably the most compelling reason for the replication" of a 3^{k-p} design "is the lack of a normal probability plot capability"? The alias scheme for a 4^{3-1} design is worked out taking advantage of equivalence to a cyclic Latin square; no mention is made of the fact that while there is a cyclic $s \times s$ Latin square for any number s of levels, for non-prime s the cyclic method quickly fails as the degree of fractionation increases. A bit is said about mixed fractions with 2, 3, and/or 4 levels. For more levels (page 280): "Fractional factorials with five levels are easier to deal with than fractional factorials for six levels because 5 is a prime number whereas 6 is not." True enough, though given the author's definition of fractional factorials, one wonders what 6-level fractions this refers to (other than the 6^{3-1} based on a cyclic Latin square). The separate section on software has some good information on package capabilities.

Chapter 7 on nested designs gets the basic structural ideas down on paper, including staggered nesting, but little more. The section on estimating variance components consists of an example where an unbiased variance estimator is negative; alternative methods are not mentioned. There is no guidance on choice of nested design.

Likewise Chapter 8 on robust designs aims only to lay out the basic concepts of control and noise factors. Section 8.1 contains some justified criticism of Taguchi designs. The sections on software and further reading are the most useful. There is also an informative discussion of control$\times$noise interaction.

Similar comments apply to Chapters 9, 11, and 12 on split-plots and variations, repeated measures designs, and multiple responses, respectively. Basic ideas are stated and discussed, examples from the literature are examined and usually criticized, and the technical level is kept very low. There are many references. Split-lot designs (section 9.2) are a topic of some recent interest in the literature, but the treatment here offers little. The multi-response optimization is unusual for design books; it is mainly an analysis topic, with here less than one page devoted to design questions.

Chapter 10 on response surface methodology is relatively longer and more detailed. Considerable attention is given to both design and analysis, the latter chiefly by example. The classical designs and numerous alternatives are listed with uneven coverage of their properties. The section on comparison of designs, where some real help could have been given, comprises a single paragraph.

Chapter 13, miscellaneous design topics, is the other lengthy chapter in the book's second half. Twenty-three mostly design topics are presented in 23 separate sections. Many of these are brief, doing no more than identifying the topic and offering some references. Some however are very interesting and leave one wanting for more, especially those design techniques of the last twenty years that have found their way into practice (e.g. supersaturated designs, space-filling designs, restricted regions of operability). There are parts of chapter 13 that are fairly classified as "modern."

Neither the strengths nor the general emphasis of *Modern Experimental Design* live up to its title, the stated goals of its preface, or its cover blurbs. This is a mostly classical text in industrial design written for non-statisticians. At least partly due to an attempt to avoid any and all mathematical rigor, there are many unnecessarily long-winded explanations that cannot escape the particular example they contemplate, where a few formulas and lines of proof would be markedly clearer and far more general. There are too few examples that demonstrate the effectiveness of the designs studied, and too many that seem bent on criticizing decisions in published examples. There is precious little advice on even simple design questions such as calculations of power or confidence interval width to determine sample size. While one strength is the many references, preferable would have been less reliance on references and more explanation of the referenced techniques. Most of the material is handled much better elsewhere, the clear exception being the separate sections addressing capabilities of design softwares. This book is not recommended.

J. P. Morgan, Department of Statistics, Virginia Tech, Blacksburg, Virginia, 24061-0439 USA.
jpmorgan@vt.edu

References.

- Atkinson, A. C.; Bailey, R. A. (2001). One hundred years of the design of experiments on and off the pages of Biometrika. Biometrika 88, 53–97.
- Box, G. E. P. (1990). Must we randomize our experiment? Quality Engineering 2, 497-502.
- Bailey, R. A.; Rowley, C. A. (1987). Valid randomization. Proc. Roy. Soc. Lond. A 410, 105–124.
- Cheng, C.-S. (1995). Some projection properties of orthogonal arrays. Ann. Statist. 23, 1223–1233.
- Hinkelmann, K.; Kempthorne, O. (1994). Design and analysis of experiments. Vol. 1. Introduction to experimental design. John Wiley & Sons, Inc., New York.