Reviewer: Russell V. Lenth
University of Iowa

Design and Analysis of Experiments: Classical and Regression Approaches with **SAS**

This book is aimed primarily at advanced undergraduate and beginning graduate students. The preface indicates that it could be used as the text for a two-semester sequence in regression and design, or selected topics could be used for a one-semester graduate course in design for statistics majors, and there is more than one semester's worth for non-statistics majors. There are probably fewer regression topics than most would want for a standard regression course, however; so I question the idea of that two-semester regression-design sequence.

There are twelve chapters that follow a fairly traditional sequence. There is a chapter with review materials (and beyond) on statistical methods and regression analysis, including multiple regression, polynomial models, and orthogonal polynomials. Chapter 2 discusses one-factor designs, with modeling, ANOVA, some follow-up tests, checking model assumptions, contrasts, and regression formulations. Chapters 3 and 4 extend these materials to two-factor designs, and introduce random effects, mixed models, and expected mean squares. Chapters 5 and 6 discuss block designs, Latin and Graeco-Latin squares. Chapters 7–9 cover 2^k and 3^k designs, fractional replication, and confounding. Chapter 10 presents materials on BIB, lattice, and nested designs. Chapter 11 introduces response-surface methods and analysis of covariance. Chapter 12 is on MANOVA.

The back cover and promotional materials for this text characterize it as being "unlike other books ... it not only covers classical experimental design theory, it also explores regression approaches." Actually, lots of design books discuss regression approaches; but after examining the presentation of topics in this text, I gather this to mean that considerable attention is given to how ANOVA models can be expressed and analyzed in terms of dummy variables. This is done in considerable detail. To obtain unique parameter estimates, the book emphasizes effects coding (where one treatment level's indicator is subtracted from each of the others), and reference-cell coding (where one indicator is omitted); and often repeats examples using both codings. In all the accompanying **SAS** examples I examined, these dummy variables are inputted explicitly as part of the data. I found no discussion of how **SAS** deals with estimability, or of its Type I–IV estimable functions.

Indeed, I have to say that if you seek a text where SAS is used as a vital tool to replace the tedium of hand calculations and to aid in exploring and graphing data, you will be sorely disappointed. The role of SAS is confined to producing regression coefficients and sums of squares. Almost all examples are confined to SAS PROCs REG, ANOVA, GLM, MIXED, and PRINT (most often of all); and even then, these are used in very limited ways. For example, I could not find an instance where residuals are saved in an output dataset, or where SAS is used to produce diagnostic plots. This is especially puzzling in that there is a fair amount of emphasis on such plots and they are shown in many examples—just not in the SAS examples.

There are several other SAS topics that are notable by their absence. I saw no examples where an INFILE statement or PROC IMPORT is used to read in data, or where a SAS binary data file is created or used; all SAS examples incorporate the data in the program using DATALINES or CARDS (!) statements. I did not see an example where a design is actually produced or randomized, via PROC PLAN, FACTEX, OPTEX. I did not see a discussion of sample-size, power, or the use of PROC POWER or GLMPOWER.

Related to the lack of data-file-related example, another prominent non-feature of this book is the lack of an accompanying CD-ROM or Web site where textbook data files can be obtained. It is surprising that this can even happen in a 2009 textbook with SAS in its title.

The mathematical level is high, with plenty of notation and detailed derivations, including matrix notations and even generalized inverses. However, the statistical level is often low, with insufficient discussion of important inferential ideas. For instance, follow-up tests are presented in Section 2.9. The methods given are the LSD, the Duncan multiple-range test (presented as "another method..."), and the Tukey HSD ("yet another test ..."). At the end of the section, we have but one paragraph (four lines long) that in any way contrasts these three methods. It states that the HSD yields narrower confidence intervals than the LSD (which is false, and at odds with the numerical example just presented), more conservative than the Duncan method (true), and the latter "has a type I error rate of α for all comparisons on an experimentwise basis" (questionable, depending on meaning of "experimentwise" which is not defined). That's it; end of discussion.

The tone of this textbook is that of a very routinized approach to data analysis, focusing on obtaining numerical answers with small datasets. One of the routines evident in the SAS examples is the inputting of data in unaligned free format, with several observations per line; and the use of PROC PRINT with an accompanying SUM statement in almost all SAS examples. I have to admit that I didn't know about the SUM statement before reading this book. Somehow, I have managed to get by all these years without displaying column sums in my data printouts.

The instructional goals are spelled out clearly on the back cover: "the author uses both manual methods and SAS programs to carry out analyses." It is important to take this very literally. There is considerable emphasis (and no relief from) hand calculations, with *additional* materials on SAS. That is, SAS programming is not treated as an integral tool in statistical analysis; it is a source of additional numerical examples and end-of-chapter exercises, and really not much more. This is very far from what I had expected (or hoped) to see. I go back far enough to remember the original version of SAS, and using PROC REGR before it was supplanted by PROC GLM. Nevertheless, I disagree with the use of the word "modern" in the back-cover description of this book.

Reviewer:

Russell V. Lenth
The University of Iowa
Department of Statistics and Actuarial Science
Iowa City, IA 52242, United States of America
E-mail: russell-lenth@uiowa.edu