



Journal of Statistical Software

October 2011, Volume 43, Book Review 5.

<http://www.jstatsoft.org/>

Reviewer: Ulrike Grömping
Beuth University of Applied Sciences Berlin

Optimal Experimental Design with R

Dieter Rasch, Jürgen Pilz, Rob Verdooren, Albrecht Gebhardt
Chapman & Hall/CRC, Boca Raton, FL, 2011.
ISBN 978-1-4398-1697-4. 345 pp. USD 99.95 (P).
<http://wwwu.uni-klu.ac.at/agebhard/OPDOE/>

The title for this book sounds quite comprehensive; unfortunately, it promises too much. A more expedient title would be something like *Aspects of Optimum Experimental Design and Some Extensions to R*. The book concentrates on selected areas of optimum experimental design, where the authors perceive gaps in R functionality. In addition to presenting theory- and formula-oriented discussions of these topics, the authors provide a collection of R functions – R package **OPDOE** – along with the formulae. That package is downloadable from the fourth author's homepage (see above). It should be emphasized that the book does not belong to the new Chapman & Hall/CRC R series; there I would expect books with far more familiarity with and emphasis on R software. This book, among more than 150 references, has a mere five references to R-related sources: a CRAN task view dated 2009 (no more specific timing; current version: [Grömping 2011a](#)), a PDF document from the useR! 2008 conference, and three books. Readers who were attracted by the title or the cover text will likely be disappointed by the book's concept.

This said, I will now discuss the book itself and some aspects of the accompanying R package. After preface and introduction, the book comes in three parts. An appendix provides more technical mathematical background like groups, fields, and finite geometries or difference sets. The first half of the book consists of Part I, which covers sample size determination for many different scenarios (Chapters 2–5), including sequential designs, for which there are also some analysis features. This part is more or less constructed like a catalogue, giving formulae – and in this part also mostly R functions – for sample sizes. It covers inference problems in completely randomized designs, various types of one- to three-way analysis of variance situations, simple regression situations and sequential designs. This is the most easily applicable part of the book. The additional R functions will come in useful.

Part II of the book (approximately a quarter of the pages) discusses balanced incomplete block designs, fractional factorial designs, and optimal designs for some regression situations with fixed regressors (Chapters 6–8). This part puts more emphasis on theory; it gives fewer examples and less code than the first part. It is dominated by the chapter on balanced incomplete block designs (BIBDs), which discusses many different methods for creating such

designs. Function `bibd` appears to be a useful addition to existing creation of BIBDs in R (mainly functions `find.BIB` and `isGYD` from package `crossdes` by Sailer 2008); I say ‘appears’, since function `bibd` is not yet included in the downloadable version of R package `OPDOE` (version 1.0-3, as of 2011-10-16). According to the book, function `bibd` can determine the smallest balanced incomplete block design for a specified number of treatments and a given block size. This is an improvement over current R functionality in package `crossdes`: there, users have to check for existence of a BIBD for specified numbers of treatments and blocks and a given block size, until they find a successful combination. Chapter 7 on fractional factorial designs is very theoretical and discusses how to obtain a complete listing of generating contrasts only, without giving explicit commands for constructing an actual design. No relevant additions to R are made in this chapter, and the existing R functionality is not properly explained or exploited (see also below). Chapter 8 again adds one useful R function (`design.reg.polynom`) and mentions one or more further such function(s) (‘the R program’) with features that seem to be a declaration of intent for future versions of R package `OPDOE`.

The final Part III (less than ten percent of the pages) briefly covers second order designs (Chapter 9) and mixture designs (Chapter 10). This part is the most unfinished. In Chapter 9 the authors made unfortunate choices especially with respect to what is covered in terms of R code: A function for central composite designs is provided, although they are already available in R (Lenth 2009; Grömping 2011b, based on the former). On the other hand, Doehlert designs are given in paper examples only; the book gives a construction principle for two-factor designs only and does not provide an R function. Also, the section on D -optimal central composite designs is not practically relevant, and the principally useful comparison of designs with respect to D -optimality lacks structure. Chapter 10 on mixture designs discusses (in Section 10.6) another unspecified R function (‘the R program’) without giving an example or a function name, again presumably a declaration of intent for package `OPDOE`; the existing functions `gen.mixture` from package `AlgDesign` (Wheeler 2011) or `mixDesign` from package `qualityTools` (Roth 2011) are not mentioned.

The R package `OPDOE` provides a collection of functions related to formulae in the book. Some of these are convenience functions, like `size.t.test` or `delta.t.test`, others can do substantial work, like `size_c.three_way_mixed_ab_in_c.model_5_a` or `bibd`. The sample size functions have logically-structured but very inconvenient names. In a recent conference talk, Albrecht Gebhardt announced a wrapper function that handles different cases through options; this will make the package more user-friendly. The package is still under development also in other respects: absence of function `bibd` and various unnamed functions was already mentioned. Online help is also still incomplete; for those functions with online help already available, users are referred to the book for details, which is not unheard of but not very user-friendly. The book’s preface announces the book as the package’s manual. This is reasonable in the sense that package functions use the same notation as the book’s formulae. However, the book falls short of all usual expectations regarding a software manual in many respects; it contains too little technical detail and too little systematic overview over the package functionality. Hopefully, the authors will eventually produce all functions mentioned or alluded to in the book, produce a usable online documentation and upload the package to the Comprehensive R Archive Network (CRAN).

At this point, I would like to comment on the lack of research into and reference to existing R functionality. Given that the book set out to fill gaps within R, I would have expected that the connection to existing software would be made explicit. This is not done; the most

striking omissions are that of package **AlgDesign** from the chapters on optimal designs (e.g., function `optFederov`) and mixture designs (function `gen.mixture`), that of package **crossdes** from the chapter on BIBDs, that of package **FrF2** (Grömping 2011c) from the chapter on fractional factorials, and that of package **rsm** (Lenth 2009) from the chapter on second order response surface designs. In the latter case, the authors re-invented the wheel with their function `design.centralcomposite`, which does not seem to be able to do more than is already available in R. Software is generally not cited among the references, not even R itself (R Development Core Team 2011) or those R packages on which the R package **OPDOE** depends, like package **crossdes**. Very few related R packages are at least mentioned in the text: package **agricolae** (de Mendiburu 2010) is criticized for its narrow scope of functionality regarding creation of balanced incomplete block designs (Sections 1.5 and 6); package **conf.design** (Venables 2010) is used in Chapter 7 for investigating the confounding structure of symmetric fractional factorial 2-level and 3-level plans; that chapter also mentions packages **faraway** (Faraway 2011, I did not find any such functionality in that package) and **AlgDesign** (misspelled as `alg.design`, Wheeler 2011) for constructing fractional factorial 2-level plans, but misses out on packages **BHH2** (Barrios 2009, function `ffDesMatrix`) or **FrF2** (function `FrF2`).

Apart from *severe* first edition issues of being increasingly unfinished after Chapter 5, the weakest point of this book is the discrepancy between label and content. Where the book can profit from the authors' profound expertise (e.g., Rasch and Herrendörfer 1986; Rasch, Verdooren, and Gowers 2007), particularly with respect to sample size and BIBDs (i.e., Chapters 1–6), book and R package provide a collection of material, which will be a welcome addition to the toolbox for planning data collection with R, once the software is in reasonable shape. Viewing sample size determination as part of optimum experimental design certainly has its merits and may have been neglected in the R perspective on the topic so far. Overall, I think that mathematically apt readers, who want to do sample size determination for non-trivial experimental setups or are interested in the ins and outs of balanced incomplete block designs, will benefit most from the book.

References

- Barrios E (2009). *BHH2: Useful Functions for Box, Hunter and Hunter II*. R package version 1.0.3, URL <http://CRAN.R-project.org/package=BHH2>.
- de Mendiburu F (2010). *agricolae: Statistical Procedures for Agricultural Research*. R package version 1.0-9, URL <http://CRAN.R-project.org/package=agricolae>.
- Faraway J (2011). *faraway: Functions and Datasets for Books by Julian Faraway*. R package version 1.0.5, URL <http://CRAN.R-project.org/package=faraway>.
- Grömping U (2011a). “CRAN Task View: Design of Experiments (DoE) & Analysis of Experimental Data.” Version 2011-08-10, URL <http://CRAN.R-project.org/view=ExperimentalDesign>.
- Grömping U (2011b). *DoE.wrapper: Wrapper Package for Design of Experiments Functionality*. R package version 0.8-4, URL <http://CRAN.R-project.org/package=DoE.wrapper>.

- Grömping U (2011c). **FrF2**: *Fractional Factorial Designs with 2-Level Factors*. R package version 1.2-8, URL <http://CRAN.R-project.org/package=FrF2>.
- Lenth RV (2009). “Response-Surface Methods in R, Using **rsm**.” *Journal of Statistical Software*, **32**(7), 1–17. URL <http://www.jstatsoft.org/v32/i07/>.
- Rasch D, Herrendörfer G (1986). *Experimental Design: Sample Size Determination and Block Designs*. Reidel Publishing Company, Dordrecht.
- Rasch D, Verdooren LR, Gowers JI (2007). *The Design and Analysis of Experiments and Surveys*. 2nd edition. Oldenbourg Verlag, München.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Roth T (2011). **qualityTools**: *Statistical Methods for Quality Science*. R package version 1.49, URL <http://CRAN.R-project.org/package=qualityTools>.
- Sailer MO (2008). **crossdes**: *Design and Randomization in Crossover Studies*. R package version 1.0-9, URL <http://CRAN.R-project.org/package=crossdes>.
- Venables WN (2010). **conf.design**: *Construction of Factorial Designs*. R package version 1.01, URL <http://CRAN.R-project.org/package=conf.design>.
- Wheeler B (2011). **AlgDesign**: *Algorithmic Experimental Design*. R package version 1.1-7, URL <http://CRAN.R-project.org/package=AlgDesign>.

Reviewer:

Ulrike Grömping
Beuth University of Applied Sciences Berlin
Department II
13353 Berlin, Germany
E-mail: groemping@bht-berlin.de
URL: <http://prof.beuth-hochschule.de/groemping/>