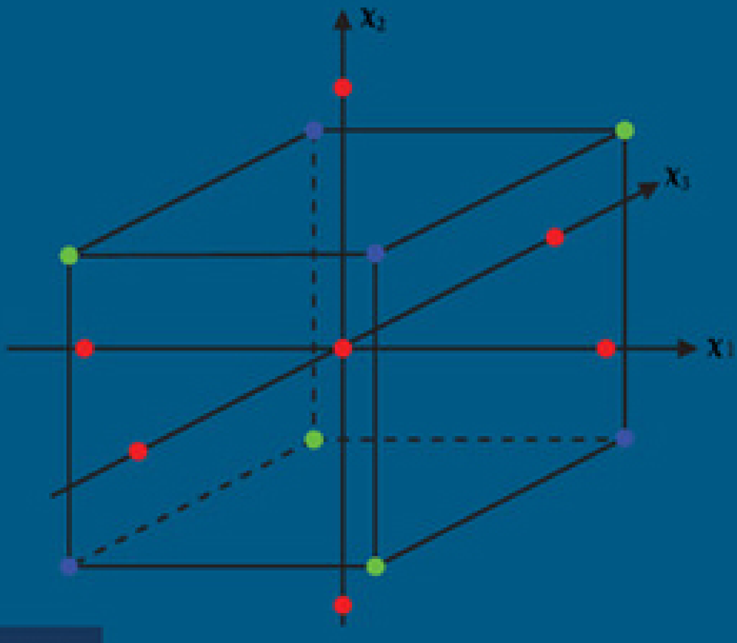
Cambridge Series in Statistical and Probabilistic Mathematics

Statistical Principles for the Design of Experiments

Applications to Real Experiments

R. Mead, S. G. Gilmour and A. Mead



CAMBRIDGE

Statistical Principles for the Design of Experiments

This book is about the statistical principles behind the design of effective experiments and focuses on the practical needs of applied statisticians and experimenters engaged in design, implementation and analysis. Emphasising the logical principles of statistical design, rather than mathematical calculation, the authors demonstrate how all available information can be used to extract the clearest answers to many questions. The principles are illustrated with a wide range of examples drawn from real experiments in medicine, industry, agriculture and many experimental disciplines. Numerous exercises are given to help the reader practise techniques and to appreciate the difference that good design can make to an experimental research project.

Based on Roger Mead's excellent *Design of Experiments*, this new edition is thoroughly revised and updated to include modern methods relevant to applications in industry, engineering and modern biology. It also contains seven new chapters on contemporary topics, including restricted randomisation and fractional replication.

R. MEAD is Emeritus Professor of Applied Statistics at the University of Reading.

S. G. GILMOUR is Professor of Statistics in the Southampton Statistical Sciences Research Institute at the University of Southampton.

A. MEAD is Senior Teaching Fellow in the School of Life Sciences at the University of Warwick.

CAMBRIDGE SERIES IN STATISTICAL AND PROBABILISTIC MATHEMATICS

This series of high-quality upper-division textbooks and expository monographs covers all aspects of stochastic applicable mathematics. The topics range from pure and applied statistics to probability theory, operations research, optimisation and mathematical programming. The books contain clear presentations of new developments in the field and also of the state of the art in classical methods. While emphasising rigorous treatment of theoretical methods, the books also contain applications and discussions of new techniques made possible by advances in computational practice.

A complete list of books in the series can be found at www.cambridge.org/statistics.
Recent titles include the following:

10. *Data Analysis and Graphics Using R,* by John Maindonald and John Braun
11. *Statistical Models,* by A. C. Davison
12. *Semiparametric Regression,* by David Ruppert, M. P. Wand and R. J. Carroll
13. *Exercises in Probability,* by Loïc Chaumont and Marc Yor
14. *Statistical Analysis of Stochastic Processes in Time,* by J. K. Lindsey
15. *Measure Theory and Filtering,* by Lakhdar Aggoun and Robert Elliott
16. *Essentials of Statistical Inference,* by G. A. Young and R. L. Smith
17. *Elements of Distribution Theory,* by Thomas A. Severini
18. *Statistical Mechanics of Disordered Systems,* by Anton Bovier
19. *The Coordinate-Free Approach to Linear Models,* by Michael J. Wichura
20. *Random Graph Dynamics,* by Rick Durrett
21. *Networks,* by Peter Whittle
22. *Saddlepoint Approximations with Applications,* by Ronald W. Butler
23. *Applied Asymptotics,* by A. R. Brazzale, A. C. Davison and N. Reid
24. *Random Networks for Communication,* by Massimo Franceschetti and Ronald Meester
25. *Design of Comparative Experiments,* by R. A. Bailey
26. *Symmetry Studies,* by Marlos A. G. Viana
27. *Model Selection and Model Averaging,* by Gerda Claeskens and Nils Lid Hjort
28. *Bayesian Nonparametrics,* edited by Nils Lid Hjort *et al*.
29. *From Finite Sample to Asymptotic Methods in Statistics,* by Pranab K. Sen, Julio M. Singer and Antonio C. Pedrosa de Lima
30. *Brownian Motion,* by Peter Mörters and Yuval Peres
31. *Probability,* by Rick Durrett
32. *Analysis of Multivariate and High-Dimensional Data,* by Inge Koch
33. *Stochastic Processes,* by Richard F. Bass
34. *Structured Regression for Categorical Data,* by Gerhard Tutz
35. *Exercises in Probability (Second Edition),* by Loïc Chaumont and Marc Yor

Statistical Principles for the Design of Experiments

R. Mead
University of Reading

S. G. Gilmour
University of Southampton

A. Mead
University of Warwick

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521862141

First published 2012

Printed and bound in the United Kingdom by the MPG Books Group

A catalogue record for this publication is available from the British Library
Library of Congress Cataloguing in Publication data
Mead, R. (Roger)
Statistical principles for the design of experiments : applications to real experiments / R. Mead, University of Reading, S.G. Gilmour, University of Southampton, A. Mead, University of Warwick.
pages cm. – (Cambridge series in statistical and probabilistic mathematics)
Includes bibliographical references and index.
ISBN 978-0-521-86214-1 (hardback)
1. Experimental design. I. Gilmour, S. G. II. Mead, A. (Andrew) III. Title.
QA279.M38825 2012
001.4′34 – dc23 2012023181

ISBN 978-0-521-86214-1 Hardback

Contents

Preface

Our aim in this book is to explain and illustrate the fundamental statistical concepts required for designing efficient experiments to answer real questions. This book has evolved from a previous book written by the first author. That book was based on 25 years of experience of designing experiments for research scientists and of teaching the concepts of statistical design both to statisticians and to experimenters. The present book is based on approximately a combined 100 years of experience of designing experiments for research scientists, and of teaching the concepts of statistical design both to statisticians and to experimenters.

The development of statistical philosophy about the design of experiments has always been dominated by mathematical theory. In contrast the influence of the availability of vastly improved computing facilities on teaching, textbooks and, most crucially, practical experimentation has been relatively small. The existence of statistical programs capable of analysing the results from any designed experiment does not imply any changes in the main statistical concepts of design. However, developments from these concepts have often been restricted by the earlier need to develop mathematical theory for design in such a way that the results from the designs could be analysed without recourse to computers. The fundamental concepts continually require reexamination and reinterpretation outside the limits implied by classical mathematical theory so that the full range of design possibilities may be considered. The result of the revolution in computing facilities is that the design of experiments should become a much wider and more exciting subject. We hope that this book will display that breadth and excitement.

The original development of the earlier book was particularly motivated by teaching postgraduate students specialising in statistics. However, the intention of this book is to reach a much wider audience. Understanding the fundamental concepts of design is essential for everybody involved in programmes of research and experimentation. In addition to this general need for an understanding of the philosophy of designing experiments there are particular aspects of design, such as the definition of experimental units (Chapter 5) or levels of replication (Chapter 6), which are relevant to virtually all research disciplines in which experimentation is required. Because of the concentration on basic concepts and their implications the book could be used for courses for final-year undergraduates provided such courses allow sufficient time for the concepts to be thoroughly discussed.

Parts of the book could also be used as text support for various tertiary education courses or short courses. Thus a course on linear models with an emphasis on data from designed studies could use Chapters 4, 9 and 18, with some examples taken from other chapters. An introductory course on the design of experiments could use Chapters 1, 2 and 3, followed by some small initial parts of Chapters 7, 8, 13 and 16. A course for final-year undergraduates

could include Chapters 5 and 6, followed by large parts of Chapters 7, 8, 10, 12, 13, 15 and 16. It would even be possible to construct a course on advanced design by including the later parts of Chapters 7 and 8, with Chapters 11, 13, 17, and some parts of 19 and 20, before presenting Chapter 0.

This book concentrates on the ideas of design rather than those of analysis, on the statistical concepts rather than the mathematical theory and on designing practically useful experiments rather than on developing classes of possible design structures. Obviously it is also necessary to consider how the data from designed experiments will be analysed and the philosophy and methods of analysis are discussed in the introductory first part of the book. There is a further chapter, 9, concerned with the analysis of data from experiments with multiple levels of information, once the ideas of designs with information at multiple levels have been considered in Chapters 7 and 8. Of course, examples of analysis punctuate many later chapters of the book. However, in all the development of design ideas, it is assumed that the analysis of data from designed experiments is not difficult when good designs and modern computing facilities are used. Consequently ideas of analysis are introduced only when they illuminate or motivate the design concepts.

The formal language of statistics is mathematical. Thus it is not possible to discuss the design of experiments without some mathematically complex formulation of models and ideas. Some of the mathematical language used in the book requires a sound mathematical background beyond school level. However, in all parts of the book it is the statistical concepts which are important, and the structure of the book hopefully allows the less-mathematical reader to bypass the more complex mathematical details. Throughout the book the development of concepts relies on many examples. We hope that readers will consider the detailed arguments of these examples. By trying to solve the problems which underlie the examples before reading through the explanation of the solutions, we believe that readers will start to develop the intuitive understanding of design concepts which is essential to good design. For the mathematically sophisticated reader the mathematical details provide additional support for the statistical concepts.

Most importantly, the book is intended to show how practical problems of designing real experiments should be solved. To stimulate this practical emphasis real examples of design problems are described at the beginning of most chapters of the book. The final chapter of the book attempts an overall view of the problem-solving aspects of design.

The areas of application used in examples in this book inevitably reflect the personal experience of the three authors. The earlier book had a strong bias towards agricultural experimentation, but we believe that that bias is much reduced in this new book and that the examples and the approach to design is much broader than previously. Whether or not that belief is justified, all the examples are intended to illustrate particular forms of problem that will be relevant in many fields of application. We hope that statisticians and research scientists in a wide range of experimental disciplines will be able to interpret and adapt the concepts discussed in the book to their own requirements through the use of analogy when the examples discussed are not directly relevant to their discipline.

For those readers familiar with the earlier book it may be helpful to explain where this new book is clearly different. Compared to the previous book, there are several new chapters, 5 on 'Experimental units', 9 on 'Multilevel analysis', 11 on 'Restricted randomisation', 14 on 'Fractional replicates', 19 on 'Multiple experiments and new variation' and 20 on 'Sequential

aspects of experiments'. Some chapters in the earlier book have disappeared, some of their material being included in a shortened form in the new book; these include 'Covariance' and 'Computer analysis programs' as sections in Chapter 4, and 'Mathematical theory for confounding and fractional replication', some parts of which appear in the new Chapters 14 and 15. The chapter on model assumptions and general models has been omitted, being not strictly relevant to a book on designing experiments.

The book is divided into an overture and two main subjects with a final coda to bring together all the previous material. Chapters 1–4 constitute the overture, providing a general introduction and the basic theory necessary for analysis of experimental data. Chapters 1, 2 and 3 should be familiar to readers who have taken an elementary course in the design of experiments. Alternatively, for those readers without any previous training in this topic, these three chapters provide an introductory presentation of the two most important ideas, blocking and factorial structure. Chapter 4 is the mathematically heavy chapter, providing the necessary theory for general linear models and the analysis of data from designed experiments, with an initial explanation of the important results at a rather simpler level. Chapter 4 also explores the universal use of computer programs and packages for the analysis of data from designed experiments and reflects on the implications of this for designing experiments.

The first main subject is unit variation and control. In Chapter 5 we examine the concept of an experimental unit, and the diverse forms of experimental units. The fundamental concepts of replication, and blocking (with either one or two systems of control) are developed in depth in Chapters 6, 7 and 8. The ideas of randomisation and restricted randomisation are explored thoroughly in Chapters 10 and 11. Our aim throughout these chapters is to distinguish the purposes and practical relevance of each concept and to eliminate the confusion about these concepts which seems to be common in the minds of many of those needing to design effective experiments. The need for analysis methods and computer programs to cope with data from designed experiments with variation occurring at more than one level is explored in Chapter 9 of this part.

The second main subject is treatment questions and structure. Chapter 12 presents an overview of the need for statisticians to be involved in all stages of discussions about the choice of treatments and the interpretation of results. The classical ideas of factorial structure and multiple and single replicates are presented in Chapter 13. In Chapter 14 we explore the uses of fractions of full factorial structures. Chapter 15 examines the combination of factorial structures with the need for units to be grouped into relatively small blocks. The choice of experimental treatments for the investigation of the responses to quantitative factors is discussed in Chapters 16 (mainly for single factor response functions) and 17 (for multifactorial response surfaces). In Chapter 18 we investigate a variety of design approaches in which different levels of experimental units are deliberately used to investigate different sets of treatment factors.

Finally in the coda, Chapter 19 explores the use of sets of experiments, often in multiple locations or at different times, and also the deliberate introduction of additional variation in experimental programmes. Chapter 20 explores various sequential aspects of experiments and experimental programmes, in particular focussing on the concept that individual experiments do not exist in isolation from past and future experimentation. Finally Chapter 0 seeks to draw the concepts of the two main subjects together to provide guidance on designing effective experiments to satisfy particular practical requirements. A book on the practical design of

experiments should start with the approach of Chapter 0 but this chapter requires knowledge from the previous chapters before it can be read and understood. Hence the number and position of this chapter.

We owe a considerable debt to many consultees and collaborators, both for the stimulus to consider why the problems they presented should be covered by a book on designing experiments and also for the many examples they have provided. There are too many for us to thank them individually here for their stimulating requests and they therefore remain anonymous (some should prefer it that way, and others are too distant in the mists of time for anything else). We have also benefitted from many discussions with colleagues at Reading and elsewhere and particularly wish to thank Richard Coe, Robert Curnow, John Fenlon, Geoff Freeman, Peter Goos, Derek Pike, Roger Stern and Luzia Trinca, without whom this book would have had a more stunted growth.

Part I

Overture

1

Introduction

1.1 Why a statistical theory of design?

The need to develop statistical theory for designing experiments stems, like the need for statistical analysis of numerical information, from the inherent variability of experimental results. In the physical sciences, this variability is frequently small and, when thinking of experiments at school in physics and chemistry, it is usual to think of 'the correct result' from an experiment. However, practical experience of such experiments makes it obvious that the results are, to a limited extent, variable, this variation arising as much from the complexities of the measurement procedure as from the inherent variability of experimental material. As the complexity of the experiment increases, and the differences of interest become relatively smaller, then the precision of the experiment becomes more important. An important area of experimentation within the physical sciences, where precision of results and hence the statistical design of experiments is important, is the optimisation and control of industrial chemical processes.

Whereas the physical sciences are thought of as exact, it is quite obvious that biological sciences are not. Most experiments on plants or animals use many plants or animals because it is clear that the variation between plants, or between animals, is very large. It is impossible, for example, to predict quantitatively the exact characteristics of one plant from the corresponding characteristics of another plant of the same species, age and origin.

Thus, no medical research worker would make confident claims for the efficacy of a new drug merely because a single patient responded well to the drug. In the field of market research, no newspaper would publish an opinion poll based on interviews with only two people, but would require a sample of at least 500, together with information about the method of selection of the sample. In a drug trial, the sample of patients would often be quite small, possibly between 20 and 100, but for a final trial before the release of the drug, as many as 2000–3000 patients might be used to detect unexpected side effects of the drug. In psychological experiments, the number of subjects used might be only 8–12. In agricultural experiments, there may be 20–100 plots of land, each with a crop grown on it. In a laboratory experiment hundreds of plants may be treated and examined individually. Or just six cows may be examined while undergoing various diets, with measurements taken frequently and in great detail.

The size of an experiment will vary according to the type of experimental method and the objective of the experiment. One of the important statistical ideas of experimental design is the choice of the size of an experiment. Another is the control of the use of experimental material. It is of little value to use large numbers of patients in the comparison of two drugs,

if all the patients given one drug are male, aged between 20 and 30, and all the patients given the other drug are female, aged 50–65. Any reasonably sceptical person would doubt claims made about the relative merits of the two drugs from such a trial. This example may seem trivially obvious, but the scientific literature in medicine and many other disciplines shows that many examples of badly planned (or unplanned) experiments occur.

And this is just the beginning of statistical design theory. From avoiding foolish experiments, we can go on to plan improvements in precision for experiments. We can consider the choice of experiments as part of research strategy and can, for example, discuss the relative merits of many small experiments or a few large experiments. We can consider how to design experiments when our experimental material is generally heterogeneous, but includes groups of similar experimental units. Thus, if we are considering the effects of applying different chemicals on the properties of different geological materials, then these may be influenced by the environment from which they are taken, as well as by the chemical treatment applied. However, we may have only two or three samples from some environments, but as many as ten samples from other environments; how then do we decide which chemicals to apply to different samples so that we can compare six different chemical treatments?

1.2 History, computers and mathematics

If we consider the history of statistical experimental design, then most of the developments have been in biological disciplines, in particular in agriculture, and also in medicine and psychology. There is therefore an inevitable agricultural bias to any discussion of experimental design. Many of the important principles of experimental design were developed in the 1920s and 1930s, in particular by R. A. Fisher. The practical manifestation of these principles was very much influenced by the calculating capacity then available. Had the computational facilities which we now enjoy been available when the main theory of experimental design was being developed it is quite possible that the whole subject of design would have developed very differently. Whether or not this belief is valid, it is certainly true that a view of experimental design today must differ from that of the 1930s, or even the 1970s. The principles have not changed, but the principles are often forgotten, and only the practical manifestation of the principles retained; these practical applications do require rethinking.

The influence of the computer is one stimulus to reassessing experimental design. Another cause for concern in the development of experimental design is the tendency for increasingly formal mathematical ideas to supplant the statistical ideas. Thus the fact that a particularly elegant piece of mathematics can be used to demonstrate the existence of groups of designs, allocating treatments to blocks of units in a particular way, begs the statistical question of whether such designs would ever be practically useful.

Although our backgrounds have been in mathematics, we believe that the presentation of statistical design theory has tended to be quite unnecessarily mathematical, and we will hope to demonstrate the important ideas of statistical design without excessive mathematical encumbrance. The language of statistical theory, like that of physics, is mathematical and there will be sections of the book where those with a mathematical education beyond school level will find a use for their mathematical expertise. However, even in these sections, which we believe should be included because they will improve the understanding of statistical