

# Cambridge Series in Statistical and Probabilistic Mathematics

|   |   |   |   |   |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 3 | 4 | 5 | 1 | 2 |
| 5 | 1 | 2 | 3 | 4 |
| 2 | 3 | 4 | 5 | 1 |
| 4 | 5 | 1 | 2 | 3 |

## Design of Comparative Experiments

R. A. Bailey

# Design of Comparative Experiments

R. A. Bailey

*Queen Mary, University of London*



**CAMBRIDGE**  
UNIVERSITY PRESS

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

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](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521683579](http://www.cambridge.org/9780521683579)

© R. A. Bailey 2008

This publication is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 2008

Printed in the United Kingdom at the University Press, Cambridge

*A catalogue record for this publication is available from the British Library*

*Library of Congress Cataloguing in Publication data*

ISBN 978-0-521-86506-7 hardback

ISBN 978-0-521-68357-9 paperback

## Design of Comparative Experiments

This book should be on the shelf of every practising statistician who designs experiments.

Good design considers units and treatments first, and *then* allocates treatments to units. It does not choose from a menu of named designs. This approach requires a notation for units that does not depend on the treatments applied. Most structure on the set of observational units, or on the set of treatments, can be defined by factors. This book develops a coherent framework for thinking about factors and their relationships, including the use of Hasse diagrams. These are used to elucidate structure, calculate degrees of freedom and allocate treatment subspaces to appropriate strata. Based on a one-term course the author has taught since 1989, the book is ideal for advanced undergraduate and beginning graduate courses. Examples, exercises and discussion questions are drawn from a wide range of real applications: from drug development, to agriculture, to manufacturing.

R. A. BAILEY has been Professor of Statistics at Queen Mary, University of London since 1994. She is a fellow of the Institute of Mathematical Statistics and a past president of the International Biometric Society, British Region. This book reflects her extensive experience teaching design of experiments and advising on its application. Her book *Association Schemes* was published by Cambridge University Press in 2004.

CAMBRIDGE SERIES IN STATISTICAL AND  
PROBABILISTIC MATHEMATICS

---

*Editorial Board*

- R. Gill (Department of Mathematics, Utrecht University)  
B. D. Ripley (Department of Statistics, University of Oxford)  
S. Ross (Department of Industrial and Systems Engineering, University of Southern California)  
B. W. Silverman (St. Peter's College, Oxford)  
M. Stein (Department of Statistics, University of Chicago)

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, optimization, 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 emphasizing rigorous treatment of theoretical methods, the books also contain applications and discussions of new techniques made possible by advances in computational practice.

*Already published*

1. *Bootstrap Methods and Their Application*, by A. C. Davison and D. V. Hinkley
2. *Markov Chains*, by J. Norris
3. *Asymptotic Statistics*, by A. W. van der Vaart
4. *Wavelet Methods for Time Series Analysis*, by Donald B. Percival and Andrew T. Walden
5. *Bayesian Methods*, by Thomas Leonard and John S. J. Hsu
6. *Empirical Processes in M-Estimation*, by Sara van de Geer
7. *Numerical Methods of Statistics*, by John F. Monahan
8. *A User's Guide to Measure Theoretic Probability*, by David Pollard
9. *The Estimation and Tracking of Frequency*, by B. G. Quinn and E. J. Hannan
10. *Data Analysis and Graphics using R*, by John Maindonald and John Braun
11. *Statistical Models*, by A. C. Davison
12. *Semiparametric Regression*, by D. Ruppert, M. P. Wand, R. J. Carroll
13. *Exercises in Probability*, by Loic 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
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

---

## Preface

---

This textbook on the design of experiments is intended for students in their final year of a BSc in Mathematics or Statistics in the British system or for an MSc for students with a different background. It is based on lectures that I have given in the University of London and elsewhere since 1989. I would like it to become the book on design which every working statistician has on his or her shelves.

I assume a basic background in statistics: estimation, variance, hypothesis testing, linear models. I also assume the necessary linear algebra on which these rest, including orthogonal projections and eigenspaces of symmetric matrices. However, people's exposure to these topics varies, as does the notation they use, so I summarize what is needed at various points in Chapter 2. Skim that chapter to see if you need to brush up your knowledge of the background.

My philosophy is that you should not choose an experimental design from a list of named designs. Rather, you should think about all aspects of the current experiment, and then decide how to put them together appropriately. Think about the observational units, and what structure they have before treatments are applied. Think about the number and nature of the treatments. Only then should you begin to think about the design in the sense of which treatment is allocated to which experimental unit.

To do this requires a notation for observational units that does not depend on the treatments applied. The cost is a little more notation; the gain is a lot more clarity. Writing  $Y_{24}$  for the response on the fourth unit with treatment 2 goes with a mindset that ignores randomization, that manages the experiment by treatment, and that does not see the need for blindness. I label observational units by lower-case Greek letters: thus  $Y(\omega)$  is the response on observational unit  $\omega$  and  $T(\omega)$  is the treatment on that unit. This notation merely mimics good practice in data recording, which has a row for each observational unit: three of the columns will be the one which names the units, the one which shows the treatments applied, and the one showing the responses. In this book, randomization, blindness and management by plot structure are recurring themes.

Most structure on the set of observational units, or on the set of treatments, can be defined by factors. I have developed a method for thinking about factors and their relationships, including the use of Hasse diagrams, which covers all orthogonal designs. The method uses the *infimum*  $F \wedge G$  of two factors (which almost everybody else, except Tjur [113], writes as  $F.G$ ) and the dual concept, the *supremum*  $F \vee G$ , which almost nobody else (again apart from Tjur) sees the need for, until degrees of freedom mysteriously go wrong. Everyone that I have taught this method to has reacted enthusiastically and adopted it. However, you need to have some idea of simple structure before you can appreciate the generality of this approach, which

is therefore delayed until Chapter 10.

The Hasse diagrams, and the insistence on naming observational units, are two features of this book that do not appear in most other books on the design of experiments. The third difference, which is relatively minor, is my notation for models in factorial designs. Expressions such as  $\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$  are compact ways of parametrizing several models at once, but they do encourage the fitting of inappropriate models (what Nelder has called *the neglect of marginality* in [83]). I take the view, explained in Chapter 5, that when we analyse data we first choose which model to fit and then estimate the parameters of that model; we do not need to know how to parametrize any of the models that we did not fit. Also in Chapter 5 I spell out three principles of modelling. The first two (Sum and Intersection) are often implicit, but their neglect can lead to contradictions. The third is Orthogonality: not everyone will agree with this (see Mead [77], for example), but I believe that we should aim for orthogonality between treatment factors wherever possible.

Another relatively minor difference in my approach is that my analysis-of-variance tables always include the grand mean. This is partly to make all the calculations easier, especially when using the Hasse diagram. A more important reason is to demonstrate that fitting a larger model after a smaller one (such as a complete two-way table after an additive model) is in principle no different from fitting treatment effects after removing the grand mean.

Unlike some topics in mathematics, Design of Experiments can set out its stall early. Thus Chapter 1 introduces most of the issues, and points the way forward to where in the book they are covered in more detail. Read this chapter to see if this book is for you.

Chapter 2 covers the simplest experiments: there is no structure on either the observational units or the treatments. This gives an opportunity to discuss randomization, replication and analysis of variance without extra complications, as well as to revise prerequisite knowledge.

Structure on the observational units is developed in Chapters 4 (simple blocking), 6 (row-column designs) and 8 (observational units smaller than experimental units). Structure on the treatments is developed in parallel, in two independent chapters. Chapter 5 deals with factorial treatments (crossed factors) while Chapter 3 covers control treatments and other ways of recursively splitting up the treatments (nested factors). Chapter 3 can be omitted in a short course, but there are some areas of application where Chapter 3 is more relevant than Chapter 5. The ‘mixed’ case of factorial treatments plus a control is covered in some detail in Chapters 1, 5 and 10; this occurs surprisingly often in practice, and is frequently misunderstood.

Chapter 8 deals with the situation when one or more of the treatment factors must be applied to something larger than observational units. This topic is often misunderstood in practice, as a glance at too many journals of experimental science shows. Every working statistician should be aware of the danger of false replication.

Chapters 7 and 9 are somewhat light relief from the main development, and could be omitted without making later chapters inaccessible. Chapter 7 applies the ideas so far to experiments on people; it also describes issues peculiar to such experiments. The reader who is concerned exclusively with such experiments is advised to continue with one of the more specialized texts, such as those recommended in the Further Reading. Chapter 9 takes a single combinatorial object—the Latin square—and uses it in several ways to design different types of experiment. This demonstrates that there is no such thing as a ‘Latin-square design’, or,

perhaps, that the phrase has many interpretations.

Chapter 10 is my favourite. It pulls all the preceding material together into a single general approach. Because it is so general, the proofs are more abstract than those in the earlier chapters, and you may want to omit them at the first reading.

Chapters 11–13 introduce three more advanced topics that a statistician needs to be aware of: incomplete-block designs, confounded factorial designs, and fractional factorial designs. Anyone who needs to use these techniques frequently should probably follow this with some more advanced reading on these topics: some suggestions are made in Further Reading.

Finally, Chapter 14 is a rerun of Chapter 1 in the light of what has been covered in the rest of the book. Confronted with an experiment to design, how should we think about it and what should we do?

Each chapter is followed by questions for discussion. Because so many aspects of designing an experiment have no single ‘right’ answer, I have used these discussion questions with my students rather than requiring written homework. Each student is required to lead the discussion at least once. Apart from the initial difficulty of persuading students that this is not a terrifying ordeal, this technique has worked remarkably well. Other students join in; they share ideas and offer helpful criticism. At the end, I comment on both the presenting student’s work and the general discussion, commending what is good, correcting any outright mistakes, and pointing out any important features that they have all missed. Every year a new set of students finds new nuances in these questions.

Some instructors may want to supplement the discussion questions with written homeworks. The Exercises at the end are provided for this purpose. They are less closely linked to the individual chapters than the questions for discussion.

## Acknowledgements

I should like to thank the following people. R.E. Waller taught me the basics at the Air Pollution Research Unit: careful data recording, meticulous verification at every stage, and the excitement of extracting patterns from data. H.D. Patterson taught me so much about designing experiments while I was working as a post-doctoral researcher under his guidance. D.J. Finney was bold enough to let me teach the course on Design of Experiments in the Statistics MSc at the University of Edinburgh within twelve months of my meeting the material. J.A. Nelder appointed me as a statistician at Rothamsted Experimental Station even though I had no formal statistical qualification. D.A. Preece introduced me to such practical matters as data sniffing. D.R. Cox has been gently nagging me to write this book for over twenty years.

Thanks, too, to all the scientists whose interesting experiments I have worked on and to all those statisticians who continue to bring me their interesting design problems. Many of these are named at the back of the book.

Finally, thanks to all those students, friends and colleagues, in the widest sense of the word, who have read drafts of part of the material and made helpful suggestions. Of course, all opinions and any remaining errors are my own.

R. A. Bailey  
December 2007



---

# Contents

---

|   |                |
|---|----------------|
| <b>Preface</b>                                    | <i>page xi</i> |
| <b>1 Forward look</b>                             | <b>1</b>       |
| 1.1 Stages in a statistically designed experiment | 1              |
| 1.1.1 Consultation                                | 1              |
| 1.1.2 Statistical design                          | 2              |
| 1.1.3 Data collection                             | 2              |
| 1.1.4 Data scrutiny                               | 3              |
| 1.1.5 Analysis                                    | 4              |
| 1.1.6 Interpretation                              | 5              |
| 1.2 The ideal and the reality                     | 5              |
| 1.2.1 Purpose of the experiment                   | 5              |
| 1.2.2 Replication                                 | 5              |
| 1.2.3 Local control                               | 6              |
| 1.2.4 Constraints                                 | 6              |
| 1.2.5 Choice                                      | 7              |
| 1.3 An example                                    | 7              |
| 1.4 Defining terms                                | 8              |
| 1.5 Linear model                                  | 14             |
| 1.6 Summary                                       | 15             |
| Questions for discussion                          | 16             |
| <b>2 Unstructured experiments</b>                 | <b>19</b>      |
| 2.1 Completely randomized designs                 | 19             |
| 2.2 Why and how to randomize                      | 20             |
| 2.3 The treatment subspace                        | 21             |
| 2.4 Orthogonal projection                         | 23             |
| 2.5 Linear model                                  | 24             |
| 2.6 Estimation                                    | 24             |
| 2.7 Comparison with matrix notation               | 26             |
| 2.8 Sums of squares                               | 26             |
| 2.9 Variance                                      | 28             |
| 2.10 Replication: equal or unequal?               | 30             |

|          |   |            |
|----------|---|------------|
| 2.11     | Allowing for the overall mean                         | 30         |
| 2.12     | Hypothesis testing                                    | 33         |
| 2.13     | Sufficient replication for power                      | 35         |
| 2.14     | A more general model                                  | 38         |
|          | Questions for discussion                              | 41         |
| <b>3</b> | <b>Simple treatment structure</b>                     | <b>43</b>  |
| 3.1      | Replication of control treatments                     | 43         |
| 3.2      | Comparing new treatments in the presence of a control | 44         |
| 3.3      | Other treatment groupings                             | 47         |
|          | Questions for discussion                              | 52         |
| <b>4</b> | <b>Blocking</b>                                       | <b>53</b>  |
| 4.1      | Types of block  | 53         |
| 4.1.1    | Natural discrete divisions                            | 53         |
| 4.1.2    | Continuous gradients                                  | 55         |
| 4.1.3    | Choice of blocking for trial management               | 55         |
| 4.1.4    | How and when to block                                 | 56         |
| 4.2      | Orthogonal block designs                              | 57         |
| 4.3      | Construction and randomization                        | 59         |
| 4.4      | Models for block designs                              | 59         |
| 4.5      | Analysis when blocks have fixed effects               | 61         |
| 4.6      | Analysis when blocks have random effects              | 67         |
| 4.7      | Why use blocks?                                       | 68         |
| 4.8      | Loss of power with blocking                           | 69         |
|          | Questions for discussion                              | 71         |
| <b>5</b> | <b>Factorial treatment structure</b>                  | <b>75</b>  |
| 5.1      | Treatment factors and their subspaces                 | 75         |
| 5.2      | Interaction   | 77         |
| 5.3      | Principles of expectation models                      | 84         |
| 5.4      | Decomposing the treatment subspace                    | 87         |
| 5.5      | Analysis  | 90         |
| 5.6      | Three treatment factors                               | 92         |
| 5.7      | Factorial experiments                                 | 97         |
| 5.8      | Construction and randomization of factorial designs   | 98         |
| 5.9      | Factorial treatments plus control                     | 99         |
|          | Questions for discussion                              | 99         |
| <b>6</b> | <b>Row-column designs</b>                             | <b>105</b> |
| 6.1      | Double blocking                                       | 105        |
| 6.2      | Latin squares   | 106        |
| 6.3      | Construction and randomization                        | 108        |
| 6.4      | Orthogonal subspaces                                  | 110        |
| 6.5      | Fixed row and column effects: model and analysis      | 110        |

|          |  |            |
|----------|--|------------|
| 6.6      | Random row and column effects: model and analysis      | 112        |
|          | Questions for discussion                               | 116        |
| <b>7</b> | <b>Experiments on people and animals</b>               | <b>117</b> |
| 7.1      | Introduction   | 117        |
| 7.2      | Historical controls                                    | 118        |
| 7.3      | Cross-over trials                                      | 118        |
| 7.4      | Matched pairs, matched threes, and so on               | 119        |
| 7.5      | Completely randomized designs                          | 120        |
| 7.6      | Body parts as experimental units                       | 120        |
| 7.7      | Sequential allocation to an unknown number of patients | 121        |
| 7.8      | Safeguards against bias                                | 122        |
| 7.9      | Ethical issues   | 124        |
| 7.10     | Analysis by intention to treat                         | 126        |
|          | Questions for discussion                               | 127        |
| <b>8</b> | <b>Small units inside large units</b>                  | <b>131</b> |
| 8.1      | Experimental units bigger than observational units     | 131        |
| 8.1.1    | The context  | 131        |
| 8.1.2    | Construction and randomization                         | 132        |
| 8.1.3    | Model and strata                                       | 132        |
| 8.1.4    | Analysis   | 132        |
| 8.1.5    | Hypothesis testing                                     | 135        |
| 8.1.6    | Decreasing variance                                    | 137        |
| 8.2      | Treatment factors in different strata                  | 138        |
| 8.3      | Split-plot designs                                     | 146        |
| 8.3.1    | Blocking the large units                               | 146        |
| 8.3.2    | Construction and randomization                         | 147        |
| 8.3.3    | Model and strata                                       | 148        |
| 8.3.4    | Analysis   | 149        |
| 8.3.5    | Evaluation   | 152        |
| 8.4      | The split-plot principle                               | 152        |
|          | Questions for discussion                               | 154        |
| <b>9</b> | <b>More about Latin squares</b>                        | <b>157</b> |
| 9.1      | Uses of Latin squares                                  | 157        |
| 9.1.1    | One treatment factor in a square                       | 157        |
| 9.1.2    | More general row-column designs                        | 158        |
| 9.1.3    | Two treatment factors in a block design                | 159        |
| 9.1.4    | Three treatment factors in an unblocked design         | 161        |
| 9.2      | Graeco-Latin squares                                   | 162        |
| 9.3      | Uses of Graeco-Latin squares                           | 166        |
| 9.3.1    | Superimposed design in a square                        | 166        |
| 9.3.2    | Two treatment factors in a square                      | 166        |
| 9.3.3    | Three treatment factors in a block design              | 166        |

|           |   |            |
|-----------|---|------------|
| 9.3.4     | Four treatment factors in an unblocked design | 167        |
|           | Questions for discussion                      | 167        |
| <b>10</b> | <b>The calculus of factors</b>                | <b>169</b> |
| 10.1      | Introduction                                  | 169        |
| 10.2      | Relations on factors                          | 169        |
| 10.2.1    | Factors and their classes                     | 169        |
| 10.2.2    | Aliasing                                      | 170        |
| 10.2.3    | One factor finer than another                 | 171        |
| 10.2.4    | Two special factors                           | 171        |
| 10.3      | Operations on factors                         | 171        |
| 10.3.1    | The infimum of two factors                    | 171        |
| 10.3.2    | The supremum of two factors                   | 172        |
| 10.3.3    | Uniform factors                               | 175        |
| 10.4      | Hasse diagrams                                | 175        |
| 10.5      | Subspaces defined by factors                  | 178        |
| 10.5.1    | One subspace per factor                       | 178        |
| 10.5.2    | Fitted values and crude sums of squares       | 178        |
| 10.5.3    | Relations between subspaces                   | 178        |
| 10.6      | Orthogonal factors                            | 178        |
| 10.6.1    | Definition of orthogonality                   | 178        |
| 10.6.2    | Projection matrices commute                   | 179        |
| 10.6.3    | Proportional meeting                          | 180        |
| 10.6.4    | How replication can affect orthogonality      | 181        |
| 10.6.5    | A chain of factors                            | 181        |
| 10.7      | Orthogonal decomposition                      | 182        |
| 10.7.1    | A second subspace for each factor             | 182        |
| 10.7.2    | Effects and sums of squares                   | 184        |
| 10.8      | Calculations on the Hasse diagram             | 185        |
| 10.8.1    | Degrees of freedom                            | 185        |
| 10.8.2    | Sums of squares                               | 187        |
| 10.9      | Orthogonal treatment structures               | 189        |
| 10.9.1    | Conditions on treatment factors               | 189        |
| 10.9.2    | Collections of expectation models             | 190        |
| 10.10     | Orthogonal plot structures                    | 193        |
| 10.10.1   | Conditions on plot factors                    | 193        |
| 10.10.2   | Variance and covariance                       | 194        |
| 10.10.3   | Matrix formulation                            | 195        |
| 10.10.4   | Strata  | 196        |
| 10.11     | Randomization                                 | 196        |
| 10.12     | Orthogonal designs                            | 197        |
| 10.12.1   | Desirable properties                          | 197        |
| 10.12.2   | General definition                            | 198        |
| 10.12.3   | Locating treatment subspaces                  | 198        |
| 10.12.4   | Analysis of variance                          | 200        |

|  |            |
|--|------------|
| 10.13 Further examples                               | 202        |
| Questions for discussion                             | 215        |
| <b>11 Incomplete-block designs</b>                   | <b>219</b> |
| 11.1 Introduction                                    | 219        |
| 11.2 Balance   | 219        |
| 11.3 Lattice designs                                 | 221        |
| 11.4 Randomization                                   | 223        |
| 11.5 Analysis of balanced incomplete-block designs   | 226        |
| 11.6 Efficiency                                      | 229        |
| 11.7 Analysis of lattice designs                     | 230        |
| 11.8 Optimality                                      | 233        |
| 11.9 Supplemented balance                            | 234        |
| 11.10 Row-column designs with incomplete columns     | 235        |
| Questions for discussion                             | 238        |
| <b>12 Factorial designs in incomplete blocks</b>     | <b>241</b> |
| 12.1 Confounding                                     | 241        |
| 12.2 Decomposing interactions                        | 242        |
| 12.3 Constructing designs with specified confounding | 245        |
| 12.4 Confounding more than one character             | 249        |
| 12.5 Pseudofactors for mixed numbers of levels       | 251        |
| 12.6 Analysis of single-replicate designs            | 253        |
| 12.7 Several replicates                              | 257        |
| Questions for discussion                             | 258        |
| <b>13 Fractional factorial designs</b>               | <b>259</b> |
| 13.1 Fractional replicates                           | 259        |
| 13.2 Choice of defining contrasts                    | 260        |
| 13.3 Weight  | 262        |
| 13.4 Resolution                                      | 265        |
| 13.5 Analysis of fractional replicates               | 266        |
| Questions for discussion                             | 270        |
| <b>14 Backward look</b>                              | <b>271</b> |
| 14.1 Randomization                                   | 271        |
| 14.1.1 Random sampling                               | 271        |
| 14.1.2 Random permutations of the plots              | 272        |
| 14.1.3 Random choice of plan                         | 273        |
| 14.1.4 Randomizing treatment labels                  | 273        |
| 14.1.5 Randomizing instances of each treatment       | 275        |
| 14.1.6 Random allocation to position                 | 275        |
| 14.1.7 Restricted randomization                      | 278        |
| 14.2 Factors such as time, sex, age and breed        | 279        |
| 14.3 Writing a protocol                              | 282        |

|   |  |            |
|---|--|------------|
| 14.3.1  | What is the purpose of the experiment? | 282        |
| 14.3.2  | What are the treatments?               | 282        |
| 14.3.3  | Methods                                | 283        |
| 14.3.4  | What are the experimental units?       | 283        |
| 14.3.5  | What are the observational units?      | 283        |
| 14.3.6  | What measurements are to be recorded?  | 283        |
| 14.3.7  | What is the design?                    | 283        |
| 14.3.8  | Justification for the design           | 284        |
| 14.3.9  | Randomization used                     | 284        |
| 14.3.10   | Plan                                   | 284        |
| 14.3.11   | Proposed statistical analysis          | 284        |
| 14.4  | The eight stages                       | 285        |
| 14.5  | A story                                | 286        |
|   | Questions for discussion               | 290        |
| <b>Exercises</b>                                    |  | <b>291</b> |
| <b>Sources of examples, questions and exercises</b> |  | <b>313</b> |
| <b>Further reading</b>                              |  | <b>319</b> |
| <b>References</b>                                   |  | <b>321</b> |
| <b>Index</b>  |  | <b>327</b> |

---

## Chapter 1

# Forward look

---

### 1.1 Stages in a statistically designed experiment

There are several stages in designing an experiment and carrying it out.

#### 1.1.1 Consultation

The scientist, or other investigator, comes to the statistician to ask advice on the design of the experiment. Sometimes an appointment is made; sometimes the approach is by telephone or email with the expectation of an instant answer. A fortunate statistician will already have a good working relationship with the scientist. In some cases the scientist and statistician will both view their joint work as a collaboration.

Ideally the consultation happens in plenty of time before the experiment. The statistician will have to ask questions to find out about the experiment, and the answers may not be immediately available. Then the statistician needs time to think, and to compare different possible designs. In complicated cases the statistician may need to consult other statisticians more specialized in some aspect of design.

Unfortunately, the statistician is sometimes consulted only the day before the experiment starts. What should you do then? If it is obvious that the scientist has contacted you just so that he can write 'Yes' on a form in response to the question 'Have you consulted a statistician?' then he is not worth spending time on. More commonly the scientist genuinely has no idea that statistical design takes time. In that case, ask enough questions to find out the main features of the experiment, and give a simple design that seems to answer the purpose. Impress on the scientist that this design may not be the best possible, and that you can do better if given more notice. Try to find out more about this sort of experiment so that you are better prepared the next time that this person, or one of her colleagues, comes to you.

Usually the scientist does not come with statistically precise requirements. You have to elucidate this information by careful questioning. About 90% of the statistician's input at this stage is asking questions. These have to be phrased in terms that a non-statistician can understand. Equally, you must not be shy about asking the scientist to explain technical terms from his field if they seem relevant.

If the scientist does have a preconceived idea of a 'design', it may be chosen from an artificially short list, based on lack of knowledge of what is available. Too many books and

courses give a list of three or four designs and manage to suggest that there are no others. Your job may be to persuade the scientist that a better design is available, even if it did not figure in the textbook from which she learnt statistics.

**Example 1.1 (Ladybirds)** A famous company (which I shall not name) had designed an experiment to compare a new pesticide which they had developed, a standard pesticide, and ‘no treatment’. They wanted to convince the regulatory authority (the Ministry of Agriculture, Fisheries and Foods) that their new pesticide was effective but did not harm ladybirds. I investigated the data from the experiment, and noticed that they had divided a field into three areas, applied one pesticide (or nothing) to each area, and made measurements on three samples from each area. I asked the people who had designed it what the design was. They said that it was completely randomized (see Chapter 2). I said that I could see that it was not completely randomized, because all the samples for each pesticide came from the same area of the field. They replied that it must be completely randomized because there were no blocks (see Chapter 4) and it was not a Latin square (see Chapter 6). In defence of their argument they quoted a respectable textbook which gives only these three designs.

### 1.1.2 Statistical design

Most of this book is about statistical design. The only purpose in mentioning it here is to show how it fits into the process of experimentation.

### 1.1.3 Data collection

In collaboration with the scientist, design a form for collecting the data. This should either be on squared paper, with squares large enough to write on conveniently, or use the modern electronic equivalent, a spreadsheet or a hand-held data-logger. There should be a row for each observational unit (see Section 1.4) and a column for each variable that is to be recorded. It is better if these variables are decided before the experiment is started, but always leave space to include extra information whose relevance is not known until later.

Emphasize to the scientist that all relevant data should be recorded as soon as possible. They should never be copied into a ‘neater’ format; human beings almost always make errors when copying data. Nor should they be invented later.

**Example 1.2 (Calf feeding)** In a calf-feeding trial each calf was weighed several times, once at birth and thereafter on the nearest Tuesday to certain anniversaries, such as the nearest Tuesday to its eight-week birthday. The data included all these dates, which proved to be mutually inconsistent: some were not Tuesdays and some were the wrong length of time apart. When I queried this I was told that only the birthdate was reliable: all the other dates had been written down at the end of the experiment by a temporary worker who was doing her best to follow the ‘nearest Tuesday’ rule after the event. This labour was utterly pointless. If the dates had been recorded when the calves were weighed they would have provided evidence of how closely the ‘nearest Tuesday’ rule had been followed; deducing the dates after the event could more accurately and simply have been done by the computer as part of the data analysis.

Sometimes a scientist wants to take the data from his field notebooks and reorganize them into a more logical order for the statistician’s benefit. Discourage this practice. Not only does



|         |     |         |    |
|---------|-----|---------|----|
| Plot 8  | 6   | Plot 23 | 0  |
|         | 0   |         | 0  |
|         | 7   |         | 0  |
|         | 3   |         | 0  |
|         | 6   |         | 0  |
|         | 0   |         | 0  |
|         | 4   |         | 0  |
|         | 5   |         | 28 |
|         | 6   |         | 0  |
|         | 4   |         | 0  |
| Average | 4.1 | Average | 28 |

Fig. 1.1. Data sheets with intermediate calculations in Example 1.3

it introduce copying errors; reordering the data loses valuable information such as which plots were next to each other or what was the time sequence in which measurements were made: see Example 1.5.

For similar reasons, encourage the scientist to present you with the raw data, without making intermediate calculations. The data will be going into a computer in any case, so intermediate calculations do not produce any savings and may well produce errors. The only benefit brought by intermediate calculations is a rough check that certain numbers are the correct order of magnitude.

**Example 1.3 (Leafstripe)** In an experiment on leafstripe disease in barley, one measurement was apparently the percentage of disease on each plot. A preliminary graph of the data showed one outlier far away from the rest of the data. I asked to see the data for the outlying plot, and was given a collection of pieces of paper like those shown in Figure 1.1. It transpired that the agronomist had taken a random sample of ten quadrats in each plot, had inspected 100 tillers (sideshoots) in each quadrat to see how many were infected, and averaged the ten numbers. Only the average was recorded in the ‘official’ data. For the outlying plot the agronomist rightly thought that he did not need a calculator to add nine zeros to one nonzero number, but he did forget to divide the total by 10. Once I had corrected the average value for this plot, it fell into line with the rest of the data.

Also try to persuade the scientist that data collection is too important to be delegated to junior staff, especially temporary ones. An experiment cannot be better than its data, but a surprising number of good scientists will put much effort into their science while believing that the data can take care of themselves. Unless they really feel part of the team, junior or temporary staff simply do not have the same motivation to record the data carefully, even if they are conscientious. See also Example 1.2.

1.1.4 Data scrutiny

After the experiment is done, the data sheets or data files should be sent to the statistician for analysis. Look over these as soon as possible for obvious anomalies, outliers or evidence of bad practice. Can that number really be a calf’s birthweight? Experienced statisticians