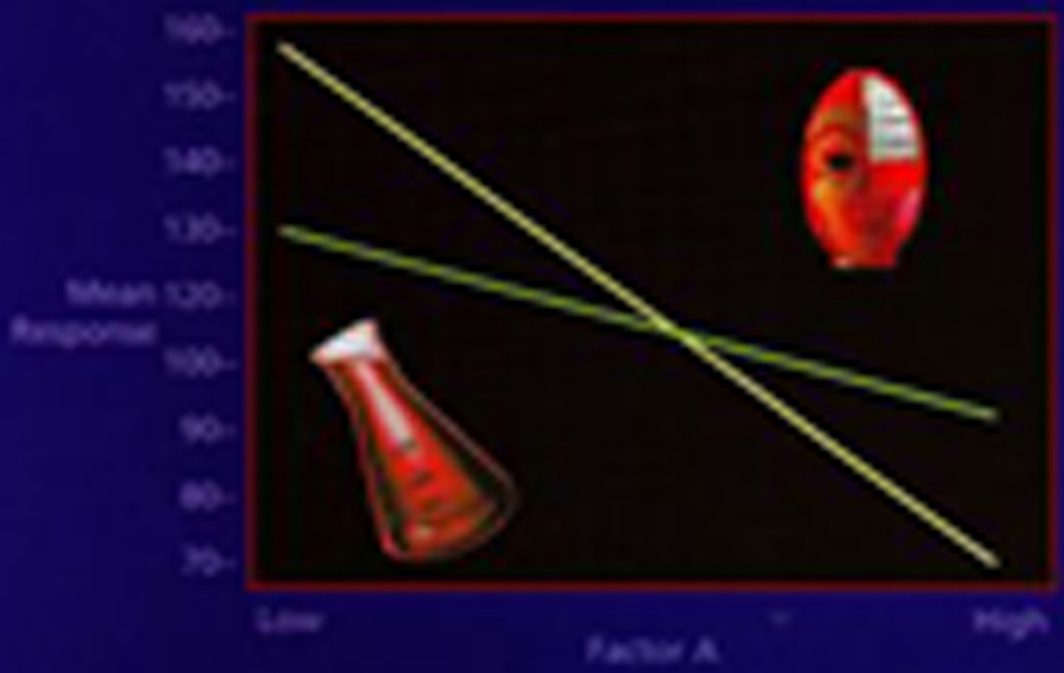
EXPERIMENTAL DESIGN TECHNIQUES IN STATISTICAL PRACTICE

A Practical Software-based Approach

W. P. Gardiner and G. Gettinby



Horwood Series in Mathematics & Applications

Experimental Design Techniques in Statistical Practice:

a practical software-based approach

"Talking of education, people have now a-days" (said he) "got a strange opinion that every thing should be taught by lectures. Now, I cannot see that lectures can do so much good as reading the books from which the lectures are taken. I know nothing that can be best taught by lectures, except where experiments are to be shewn. You may teach chymestry by lectures — You might teach making of shoes by lectures!"

James Boswell: *Life of Samuel Johnson, 1766* (1709-1784)

The direction in which education starts a man will determine his future life.

Plato: *The Republic, Book V*
(427-347 BC)

THE AUTHORS

Bill Gardiner

Dr Bill Gardiner graduated with first class honours in mathematics at the University of Strathclyde and later with a Ph.D. in statistics and mathematical modelling. He has devoted his professional life to high quality teaching of statistics to under-graduates across a range of disciplines and also to postgraduate MSc students in industrial mathematics. He has provided consultancy support for over 16 years to biologists, chemists and health providers. Much of his work has involved the use of experimental designs within practical problems which, with his teaching and consultancy experience, culminated in the writing of this book together with George Gettinby. He currently lectures at Glasgow Caledonian University, where he emphasises the importance and practical benefits of statistical data analysis as a fundamental and integral part of data interpretation for students of all disciplines. He has also recently published work on unbalanced experimental designs and two introductory statistics books for bioscientists and chemists. This present textbook covers both elementary and advanced methods and reflects Dr Gardiner's successful teaching style in practice, where the emphasis is on using software as an aid to statistical analysis and decision-making.

George Gettinby

George Gettinby is Professor and Chairman in the Department of Statistics and Modelling Science at the University of Strathclyde, Glasgow. He graduated in applied mathematics at Queens University Belfast and obtained a DPhil for his thesis on mathematical and statistical modelling at the University of Ulster. He subsequently qualified as a chartered statistician. Over the last twenty years he has been an adviser to international agencies, industry and governmental bodies on the design and analysis of studies using experimental design methods. His research and teaching interests have focused on the use of statistical and mathematical models for the study of the environment and the control of diseases. His industrial interests have centred around the research, development and manufacture of human and animal medicines. In recent years he has taught many short courses for industry that promote the use of statistical methods for assessing the quality of products. He is a member of the Royal Statistical Society, Professional Statisticians in Industry and the UK Medicines Commission. In 1997 he was elected a Fellow of the Royal Society of Edinburgh.

Experimental Design Techniques in Statistical Practice:

a practical software-based approach

William P. Gardiner, Bsc, PhD
Lecturer in Statistics
Glasgow Caledonian University

and

George Gettinby, BSc, Dphil, CStat
Professor of Statistics and Modelling Science
University of Strathclyde
Glasgow

Horwood Publishing
Chichester

First published in 1998 by
HORWOOD PUBLISHING LIMITED
International Publishers
Coll House, Westergate, Chichester, West Sussex, PO20 6QL
England

COPYRIGHT NOTICE
All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the permission of Horwood Publishing, International Publishers, Coll House, Westergate, Chichester, West Sussex, England

© W.P. Gardiner & G. Gettinby, 1998

British Library Cataloguing in Publication Data
A catalogue record of this book is available from the British Library

ISBN 1-898563-35-7

Printed and bound in Great Britain by MPG Books Ltd, Bodmin, Cornwall

Table of Contents

Authors' Preface xiii

Glossary xv

Chapter 1 Introduction
1.1 Introduction 1
1.2 Information on the Statistical Software 2
1.2.1 SAS 3
1.2.2 Minitab 4
1.3 Summarising Experimental Data 6
1.3.1 Graphical Presentations 6
1.3.2 Numerical Summaries 8
1.4 The Normal Distribution Within Data Analysis 11
1.5 Outliers in Data 12
1A Appendix: Introductory Software Information 14

Chapter 2 Inferential Data Analysis for Simple Experiments
2.1 Introduction 19
2.2 Basic Concepts of Inferential Data Analysis 19
2.3 Inference Methods for Two Sample Studies 23
2.3.1 Hypothesis Test for Difference in Mean Responses 24
2.3.2 Confidence Interval for Difference in Mean Responses 27
2.3.3 Hypothesis Test for Variability 28
2.3.4 Confidence Interval for the Ratio of Two Variances 29
2.3.5 Non-parametric Methods 30
2.4 Inference Methods for Paired Sample Studies 30
2.4.1 Hypothesis Test for Mean Difference in Responses 31
2.4.2 Confidence Interval for Mean Difference 33
2.4.3 Hypothesis Test for Variability 34
2.4.4 Non-parametric Methods 36
2.5 Sample Size Estimation in Design Planning 36
2.5.1 Sample Size Estimation for Two Sample Studies 36
2.5.2 Sample Size Estimation for Paired Sample Studies 38
2.6 Validity and Good Statistical Practice 39
2.7 Protocols 40
2.8 Commonly Occurring Design Mistakes 41
Problems 42
2A.1 Appendix: Software Information for Two Sample Studies 44
2A.2 Appendix: Software Information for Paired Sample Studies 47

Chapter 3 One Factor Designs
3.1 Introduction 51
3.2 Completely Randomised Design 53
3.2.1 Design Structure 53
3.2.2 Model for the Measured Response 54
3.2.3 Assumptions 55
3.2.4 Exploratory Data Analysis 56
3.3 ANOVA Principle for the Completely Randomised Design 57

3.3.1 Hypotheses 57
3.3.2 ANOVA Table 57
3.3.3 Treatment Test Statistic 58
3.4 Follow-up Analysis Procedures 59
3.4.1 Main Effects Plot 60
3.4.2 Standard Error Plot 60
3.4.3 Multiple Comparisons 61
3.4.4 Linear Contrasts 64
3.4.5 Orthogonal Polynomials 67
3.4.6 Treatment Effect Estimation 67
3.4.7 Model Fit 69
3.5 Diagnostic Checking of Model Assumptions 69
3.5.1 Graphical Checks 70
3.5.2 Statistical Checks 73
3.6 Power Analysis in Design Planning 74
3.6.1 Power Estimation 75
3.6.2 Sample Size Estimation 76
3.7 Non-parametric Alternative to the ANOVA Based Treatment Tests 77
3.7.1 The Kruskal-Wallis Test of Treatment Differences 77
3.7.2 Multiple Comparison Associated with the Kruskal-Wallis Test 78
3.7.3 Linear Contrasts 79
3.8 Data Transformations 80
Problems 82
3A Appendix: Software Information for Completely Randomised Design 84

Chapter 4 One Factor Blocking Designs
4.1 Introduction 89
4.2 Randomised Block Design 89
4.2.1 Design Structure 90
4.2.2 Model for the Measured Response 90
4.2.3 Assumptions 91
4.2.4 Exploratory Data Analysis 91
4.3 ANOVA Principle for the Randomised Block Design 92
4.3.1 Hypotheses 92
4.3.2 ANOVA Table 92
4.3.3 Test Statistics 93
4.4 Follow-up Analysis Procedures for Randomised Block Designs 95
4.5 Additional Aspects of Blocking Designs 97
4.5.1 Power Analysis 97
4.5.2 Missing Observations 98
4.5.3 Efficiency 98
4.6 Non-parametric Alternative to ANOVA Based Treatment Tests 98
4.6.1 Friedman Test of Treatment Differences 99
4.6.2 Multiple Comparison Associated with Friedman's Test 99
4.7 Incomplete Block Designs 100
4.7.1 Balanced Incomplete Block Design 101
4.7.2 ANOVA Principle 102
4.7.3 Follow-up Analysis Procedures 103
4.7.4 Other Incomplete Designs 104
4.8 Latin Square Design 104
4.8.1 Design Structure 104

4.8.2 Model for the Measured Response 105
4.8.3 ANOVA Principle 106
4.8.4 Follow-up Analysis Procedures 108
4.8.5 Missing Observations 110
4.8.6 Efficiency 111
Problems 111
4A.1 Appendix: Software Information for Randomised Block Design 114
4A.2 Appendix: Software Information for Balanced Incomplete Block Design 116
4A.3 Appendix: Software Information for Latin Square Design 117

Chapter 5 Factorial Experimental Designs
5.1 Introduction 119
5.2 Two Factor Factorial Design with *n* Replications per Cell 120
5.2.1 Design Structure 121
5.2.2 Model for the Measured Response 123
5.2.3 Exploratory Data Analysis 124
5.3 ANOVA Principle for the Two Factor Factorial Design 124
5.3.1 Hypotheses 125
5.3.2 ANOVA Table 125
5.3.3 Test Statistics 125
5.4 Follow-up Analysis Procedures for Factorial Designs of Model I Type 127
5.4.1 Significant Interaction 128
5.4.2 Non-significant Interaction but Significant Factor Effect 130
5.4.3 Linear Contrasts 130
5.4.4 Orthogonal Polynomials 130
5.4.5 Estimation 131
5.4.6 Diagnostic Checking 132
5.5 Overview of Data Analysis for Two Factor Factorial Designs 134
5.6 Power Analysis in Two Factor Factorial Designs 134
5.7 Non-parametric Inference for a Two Factor Factorial Design 135
5.8 Random Effects - Model II/Mixed Model Experiments 136
5.8.1 Additional Effect Assumptions 137
5.8.2 *EMS* Expressions and Test Statistic Derivation 138
5.8.3 Analysis Components for Models with Random Effects 140
5.8.4 Power Analysis for Random Effects 140
5.9 Unbalanced Two Factor Factorial Design 141
5.9.1 Type I to IV Sums of Squares 142
5.9.2 Type I to III Hypotheses 144
5.10 Three Factor Factorial Design with *n* Replications per Cell 147
5.10.1 Model for the Measured Response 148
5.10.2 ANOVA Principle and Test Statistics 148
5.10.3 Overview of Data Analysis for Three Factor Factorial Designs 152
5.10.4 Pooling of Factor Effects 155
5.10.5 Unbalanced Approaches 156
5.11 Analysis of Covariance 156
Problems 156
5A.1 Appendix: Software Information for Two Factor Factorial Designs 162
5A.2 Appendix: Software Information for Three Factor Factorial Designs 164

Chapter 6 Hierarchical Designs
6.1 Introduction 166
6.2 Two Factor Nested Design 166
6.2.1 Design Structure 166
6.2.2 Model for the Measured Response 168
6.2.3 Exploratory Data Analysis 169
6.3 ANOVA Principle for the Two Factor Nested Design 169
6.3.1 Hypotheses 169
6.3.2 ANOVA Table 170
6.3.3 Test Statistics 170
6.4 Follow-up Analysis 172
6.4.1 Analysis of Factor Effects 172
6.4.2 Mixed Model - Significant Nested Factor 173
6.4.3 Mixed Model and Model II Variability Analysis 174
6.4.4 Diagnostic Checking 175
6.5 Other Features Associated with Two Factor Nested Designs 171
6.5.1 Relative Efficiency 177
6.5.2 Power Analysis 177
6.5.3 Unequal Replicates 178
6.6 Three Factor Nested Design 178
6.6.1 Model for the Measured Response 178
6.6.2 ANOVA Table and Test Statistics 179
6.7 Repeated Measures Design 179
6.7.1 Design Structure 179
6.7.2 Model for the Measured Response 180
6.7.3 ANOVA Table and Test Statistics 182
6.7.4 Follow-up Procedures 186
6.7.5 Diagnostic Checking 188
6.7.6 Unbalanced Designs 189
6.8 CrossOver Design 190
6.8.1 Design Structure 191
6.8.2 Model for the Measured Response 192
6.8.3 ANOVA Table and Test Statistics 193
6.9 Split-Plot Designs 196
6.9.1 Design Structure 196
6.9.2 Model for the Measured Response 196
6.9.3 ANOVA Table and Test Statistics 197
Problems 198
6A.1 Appendix: Software Information for Nested Designs 202
6A.2 Appendix: Software Information for Repeated Measures Designs 204
6A.3 Appendix: Software Information for CrossOver Designs 205

Chapter 7 Two-level Factorial Designs
7.1 Introduction 207
7.2 Contrasts and Effect Estimation 209
7.2.1 Contrasts 209
7.2.2 Effect Estimation 212
7.2.3 Missing Values 214
7.3 Initial Analysis Components 214
7.3.1 Exploratory Analysis 214

7.3.2 Effect Estimate Plots 214
7.3.3 Data Plots and Summaries 218
7.4 Statistical Components of Analysis 220
7.4.1 Statistical Assessment of Proposed Model 220
7.4.2 Prediction 221
7.4.3 Diagnostic Checking 222
7.5 General Unreplicated 2^k Designs 224
7.6 Use of Replication 226
7.7 Three-level Designs 227
Problems 227
7A Appendix: Software Information for Two-level Factorial Designs 229

Chapter 8 Two-level Fractional Factorial Designs
8.1 Introduction 234
8.2 Confounding 234
8.3 Block Construction Procedures 236
8.3.1 Even/Odd Method 236
8.3.2 Linear Combination Method 238
8.4 Fractional Factorial Designs 239
8.4.1 Alias Structure 242
8.4.2 Design Resolution 245
8.5 Analysis Components for Fractional Factorial Designs 248
8.5.1 Exploratory Analysis 249
8.5.2 Effect Estimates Analysis 249
8.5.3 Data Plots and Summaries 250
8.5.4 Statistical Components 251
8.6 Three-level Fractional Factorial Designs 258
Problems 259
8A Appendix: Software Information for Two-level Fractional Factorial Designs 262

Chapter 9 Two-level Orthogonal Arrays
9.1 Introduction 267
9.2 Array Structures 267
9.3 Experimental Plans 270
9.3.1 Plans Based on the $OA_8(2^7)$ Structure 270
9.3.2 Plans Based on the $OA_{16}(2^{15})$ Structure 272
9.3.3 Aliasing and Resolution in Orthogonal Arrays 274
9.3.4 Saturated Designs 276
9.4 Analysis Components for Two-Level Orthogonal Arrays 277
9.5 Three-level Orthogonal Arrays 281
9.6 Plackett-Burman Designs 282
Problems 283
9A Appendix: Software Information for Two-Level Orthogonal Array Designs 286

Chapter 10 Taguchi Methods
10.1 Introduction 289
10.2 Design Structures 291
10.3 Performance Statistics 297
10.4 Data Analysis Components for Two-level Taguchi Parameter Designs 300

10.4.1 Estimation of Factor Effects 300
10.4.2 Factor Effect Plots 301
10.4.3 Data Plots 302
10.4.4 ANOVA Modelling 304
10.5 Data Analysis Components for Three-level Taguchi Parameter Designs 307
10.5.1 Data Plots 309
10.5.2 ANOVA Modelling 312
10.6 Other Data Analysis Components 312
10.6.1 Exponential/Weibull Plots 312
10.6.2 Test of Homogeneity of the Factor Mean Squares 313
10.7 Final Comments 313
Problems 314
10A.1 Appendix: Software Information for Two-level Taguchi Designs 317
10A.2 Appendix: Software Information for Three-level Taguchi Designs 320

Chapter 11 Response Surface Methods
11.1 Introduction 322
11.2 First-order Design Structures 323
11.2.1 Two-level Design 323
11.2.2 Simplex Design 324
11.3 Response Surface Analysis for First-order Designs 324
11.3.1 First-order Response Model 324
11.3.2 Parameter Estimation 326
11.3.3 ANOVA Table 327
11.3.4 Adequacy of Fit 328
11.3.5 Statistical Validity of Fitted Model 331
11.3.6 Additional Aspects of Analysis 331
11.3.7 Contour Plot 332
11.4 Method of Steepest Ascent 332
11.5 Second-order Design Structures 335
11.5.1 Central Composite Design 335
11.5.2 Three-level Design 338
11.5.3 Box-Behnken Design 338
11.5.4 Face-Centred Design 338
11.5.5 Blocking Design 339
11.6 Response Surface Analysis for Second-order Designs 340
11.6.1 Second-order Response Model 340
11.6.2 Parameter Estimation 341
11.6.3 Adequacy of Fit 343
11.6.4 Statistical Validity of Fitted Model 343
11.6.5 Additional Analysis 344
11.6.6 Contour Plot 345
11.6.7 Exploration of Response Surface 346
11.7 Mixture Designs 348
11.8 Further Aspects of Response Surface Methods 349
11.8.1 Multiple Responses and Non-parametric Methods 349
11.8.2 Optimal Design Theory 349
Problems 349
11A Appendix: Software Information for Response Surface Designs 352

Bibliography 355

Appendix A Statistical Tables 359
Table A.1 Critical Values of the Student's t Distribution 360
Table A.2 Critical Values of the χ^2 Distribution 361
Table A.3 Critical Values of the F Distribution 362
Table A.4 Table of z Scores for the Standard Normal Distribution 366
Table A.5 5% Critical Values for the Studentised Range Statistic 367
Table A.6 1% Critical Values for the Studentised Range Statistic 368
Table A.7 Orthogonal Polynomials 369
Table A.8 Critical Values for the Ryan-Joiner Correlation Test of Normality 370
Table A.9 Power Values for Treatment F Test Based on the Non-central $F(f_1, f_2, \lambda)$ Distribution for 5% Significance Testing 371

Appendix B Answers to Selected Problems 374

Subject Index 384

This page intentionally left blank

Authors' Preface

Statistical data analysis techniques are increasingly being relied upon to translate experimental data into useful knowledge. Familiarity with statistical techniques is necessary if scientists, industrial personnel, and researchers are to design experiments to obtain the most relevant data for the specified objectives and if they are to use the data collected to best advantage in the advancement of the knowledge of new phenomena, processes, and products.

The development of experimental design principles is generally attributed to Sir Ronald Fisher who was concerned with agricultural research in the 1920s for the improvement of yield. Further development of these initial principles was provided by innovators such as Frank Yates and George Box, most notably through their contributions to agriculture and industrial experimentation. The areas in which experimental designs can be applied have continued to expand. Recently, the increased awareness of Taguchi methods for quality improvement has led to an increase in the use of statistics and experimental design in industry, though not necessarily all industry. New areas of application continue to be developed through increased awareness of the benefits of experimental design and through improved software facilities.

This text will provide an introduction to the diverse subject area of experimental design. Despite its length, it does not attempt to be an all embracing guide to all aspects of experimental designs. There are numerous other excellent texts available for that purpose. The important principles that we instill are that stating the aims and objectives, planning, and data analysis go hand in hand and must be seen as part of the overall investigative process. Inefficient designs represent wasted effort and so ensuring that these three principles are addressed is vital to efficient design and analysis.

A particular purpose of this book is to develop in the reader an appreciation and understanding of experimental designs and to equip the reader with the ability to use experimental designs in a practical way. The book illustrates experimental designs using practical problems from a wide range of application areas. We have de-emphasised the mathematical aspects underpinning experimental designs to concentrate more on the practicality of design usage and the subsequent data handling. We believe this approach will provide the reader with a greater feel and stronger understanding of how to use experimental designs in practice. The data handling is explained from both exploratory data analysis and inferential data analysis aspects through the provision of detailed solutions. This enables the reader to develop a sound understanding of how to analyse data and of the role statistical methods can play within both the design and interpretational aspects of experimentation. We concur with the trend of including more exploratory data analysis in Statistics teaching to enable data to be explored visually and numerically for inherent patterns. This aspect of data analysis has been incorporated in all the illustrations. Each chapter also contains simple, practical, and applicable problems for the reader to attempt to provide additional illustrations of the concepts and data analysis principles described. Summary solutions to selected problems are presented at the end of the text.

The use of statistical software packages has increased markedly in recent years, with inefficient and inappropriate usage still a frequent occurrence. With this in mind, we have decided to base all aspects of the design illustrations presented on the use of software output to reflect the wide availability of statistical software for data handling. It must be appreciated, however, that software is only a tool to aid data presentation for analysis purposes. The investigator must use the output created to interpret the data with respect to the specified experimental objectives. The emphasis we have placed on software usage has enabled the calculation aspects of the data analysis procedures to be kept to a minimum. This has allowed us to focus more on describing the underlying principles and methods with which to expedite the data analysis. Software usage also enables the data to be viewed and interpreted from graphical, summary, and inferential perspectives, so providing a more comprehensive base for applicable data analysis than calculation methods alone can provide. It is this philosophy which is the basis of the practical illustrations within this book.

Numerous statistical software packages are available and it was difficult to decide which to include in the book. We chose to use both Minitab and SAS as they are simple to use, ubiquitous, and compatible in most of their operations with the operational features of Windows software. Minitab is widely available throughout educational institutions while SAS is used extensively in industry. We acknowledge that this decision may not suit everyone but the data presentation and analysis principles have been presented in a way that can be readily transferred to other software packages. Detailed information and explanations of software operation are provided in appendices at the end of each chapter for both Minitab and SAS to provide the reader with full information on how to use the software to obtain the illustrated statistical output.

Acknowledgements

We wish to express our sincere thanks to Dr Bill Byrom, Professor Stuart Reid, and Dr Xuerong Mao for reviewing the manuscript. Their many constructive suggestions and helpful criticisms improved the structure and explanations provided in the text. We also wish to thank the many journals and publishers who have graciously granted us permission to reproduce materials from their publications. Thanks are also due to our publisher, Ellis Horwood MBE, for the enthusiasm he showed for the project and the assistance he has provided in the preparation of the manuscript. Finally and most importantly, we must express our appreciation and heartfelt thanks for the support of our families, Moira, Debbie and Greg, and Ruth, Michael, and Peter, through the long running saga of text development, preparation, review, and final typing. This book is dedicated to them and we hope they have as many happy hours reading it as we did preparing it!

W. P. Gardiner
Glasgow Caledonian University
w.gardiner@gcal.ac.uk

G. Gettinby
University of Strathclyde
g.gettinby@stams.strath.ac.uk

January 1998

Glossary

Accuracy The level of agreement between replicate determinations of a measurable property and a reference or target value.

Aliasing The sharing of contrast expressions and sum of squares of separate factorial effects.

Alternative hypothesis A statement reflecting a difference or change in the level of a response as a result of experimental intervention, denoted by H_1 or H_A or AH.

Analysis of Variance (ANOVA) The technique of separating, mathematically, the total variation within experimental measurements into sources corresponding to controlled and uncontrolled components.

BIBD Balanced Incomplete Block Design.

Blocking The grouping of experimental units into homogeneous blocks to remove an extraneous source of response variation.

Boxplot A data plot comprising tails and a box from lower to upper quartile separated in the middle by the median for detecting data spread and patterning together with the presence of outliers.

CCD Central Composite Design.

CI Confidence interval, an interval or range of values which contains an unknown parameter with a specified probability.

COD CrossOver Design.

Contrasts Corresponds to linear combinations of treatments for specific treatment comparisons, also the underpinning components of two-level designs.

Confounding The design technique for blocking a factorial experiment where information on certain treatment effects is sacrificed as they are indistinguishable from the block effects.

CRD Completely Randomised Design.

CV Coefficient of variation, a dimensionless quantity which is a measure of the relative precision of replicated experimental data.

Decision rule Mechanism for using test statistic or p value to decide whether to accept or reject the null hypothesis in inferential data analysis.

Defining contrast A treatment effect confounded with the blocks or fractions in Fractional Factorial Designs.

Defining relation The complete specification of sacrificed treatment effects in a Fractional Factorial Design expressed as the set of factors equal to the identity column I.

Descriptive statistics The graphical presentation and calculation of summary statistics for experimental data.

df Degrees of freedom, number of independent measurements that are available for estimation, generally corresponds to number of measurements minus number of parameters to estimate.

Diagnostic checking An analysis tool for assessment of the assumptions associated with an inferential data analysis procedure.

Dispersion The level of variation within collected data corresponding to the way data cluster around their "centre" value.

Dotplot A data plot of recorded data where each observation is presented as a dot to display its position relative to other measurements within the data set.

EDA Exploratory data analysis, visual and numerical mechanisms for presenting and analysing data to help gain an initial insight into the structure and patterning prevalent within the data.

Error Deviation of a response measurement from its true value.

Estimation Methods of estimating the magnitude of an experimental effect within an investigation.

Experiment A planned inquiry to obtain new information on a measurable, or observable, outcome or to confirm results from previous studies.

Experimental design The experimental structure used to generate practical data for interpretative purposes.

Experimental plan Step-by-step guide to experimentation and subsequent data analysis.

Experimental unit An experimental unit is the physical experimental material to which one application of a treatment is applied, e.g. manufactured product, water sample, food specimen, subject.

FD Factorial Design.

FFD Fractional Factorial Design.

Fixed effects The treatments to be tested correspond to those specifically chosen for investigation or to the only ones associated with an investigation.

Heteroscedastic Data exhibiting non-constant variability as the mean changes.

Homoscedastic Data exhibiting constant variability as the mean changes.

Inferential data analysis Inference mechanisms for testing the statistical significance of collected data through weighing up the evidence within the data for or against a particular outcome.

Interaction The joint influence of treatment combinations on a response which cannot be explained by the sum of the individual factor effects.

Location The centre of a data set which the recorded responses tend to cluster around, e.g. mean, median.

LS Latin Square design.

Main effects Independent factor effects reflecting the change in a response as a result of changing the factor levels.

MD Mixture Design.

Mean The arithmetic average of a set of experimental measurements.

Median The middle observation of a set of experimental measurements when expressed in ascending order of magnitude.

Mixed model experiment An experiment in which factors of both fixed and random effect type appear.

Model The statistical mechanism where an experimental response is explained in terms of the factors controlled in the experiment.

Model I experiment An experiment in which the treatments, or factors, tested all correspond to fixed effects.

Model II experiment An experiment in which the treatments, or factors, tested all correspond to random effects.

Multiple comparison procedures Statistical procedures for pairwise treatment comparison which provide an understanding of how detected treatment differences are occurring.

ND Nested Design.

Non-parametric procedures Methods of inferential data analysis, often based on ranking, which do not require the assumption of normality for the measured response.

Normal (Gaussian) The most commonly applied population distribution in Statistics, the assumed distribution for a measured response in parametric inference.

Null hypothesis A statement reflecting no difference between observations and target or between sets of observations as a result of experimental intervention, denoted H_0 or *NH*.

OAD Orthogonal Array Design.

Observation A measured or observed data value from a study or an experiment.

OFAT One-factor-at-a-time experimentation.

OLS Ordinary least squares, a parameter estimation technique used within regression modelling to determine the best fitting relationship for a response Y in terms of one or more experimental variables.

Orthogonality The property of a design matrix whereby the inner product of any pair of columns is zero.

Orthogonal polynomials Specific treatment contrasts which can assess for evidence of trend effects in quantitative treatments.

Outlier A recorded response measurement which differs markedly from the majority of the data collected.

***p* Value** The probability that a calculated test statistic value could have occurred by chance alone, provides a measure of the probability that the level of treatment difference detected has occurred purely by chance, compared to significance level.

Paired sampling A design principle where experimental material to be tested is split into two equal parts with each part tested on one of two possible treatments.

Parameters The terms included within a response model which require to be estimated and assessed for their statistical significance.

Parametric procedures Methods of inferential data analysis based on the assumption that the measured response data conform to a normal distribution.

PBD Plackett-Burman Design.

Power The probability of correctly rejecting a false null hypothesis, power = 100[1 – P(Type II error)], often set at 80%.

Power analysis An important part of design planning to assess suitability of design structure for its intended purpose.

Precision The level of agreement between replicate measurements of a measurable property.

Protocol An outline of the study approach specifying objectives, sampling strategy, power analysis, and details of planned data analysis.

Quality assurance (QA) Procedures concerned with monitoring of laboratory practice, manufacturing practice, and measurement reporting to ensure quality of reported measurements.

Quality control (QC) Mechanisms for checking that reported measurements are free of error and conform to acceptable accuracy and precision.

Quantitative data Physical measurements of a study outcome conforming to a validated scale system.

Random effects The treatments to be tested represent a random sample from larger population.

Random error Causes response measurements to fall either side of a target affecting data precision.

Randomisation Reduces the risk of bias in experimental results, concerned with selection of experimental units for use within an experiment and run order of experiments.

Range A simple measure of data spread.

Ranking Ordinal number corresponding to the position of a measurement value when measurements are placed in ascending order of magnitude.

Repeatability A measure of the precision of a method expressed as the agreement attainable between independent determinations performed by a single individual using the same instrument and techniques in a short period of time.

Replication The concept of repeating experimentation to produce multiple measurements of the same response to enable data accuracy and precision to be estimated.

Reproducibility A measure of the precision of a method expressed as the agreement attainable between determinations performed in different locations.

Residuals Estimates of model error, determined as the difference between the recorded observations and the model's fitted values.

Resolution The ability of a two-level or three-level Fractional Factorial Design to provide independent factor effect estimates of the main components of interest.

Response The characteristic measured or observed in a study.

RMD Repeated Measures Design.

Robust statistics Data summaries which are unaffected by outliers and spurious measurements.

RSM Response Surface Methods.

Sample A set of representative measurements of a measurable or observable outcome.

Sample size estimation Integral part of design planning, ensure sufficient measurements are collected to enable study objectives to be properly assessed.

Screening experiments Simple to implement and analyse multi-factor experiments for the early stages of projects to identify important factors.

Significance level The probability of rejecting a true null hypothesis, P(Type I error), typically set at 5% or 0.05.

Skewness Shape measure of data for assessing lack of symmetry.

SNK Student-Newman-Keuls multiple comparison procedure.

SPC Statistical Process Control.

SPD Split-Plot Design.

SQC Statistical Quality Control.

Standard deviation A magnitude dependent measure of the absolute precision of replicate experimental data.

Systematic error Causes response measurements to be in error affecting data accuracy.

Test statistic A mathematical formulae which provides a measure of the evidence that the study data provide in respect of acceptance or rejection of the null hypothesis, numerically estimable using study data.

Taguchi methods Experimental design and analysis techniques initially pioneered by Genichi Taguchi for quality improvement of products and processes.

Transformation A technique of re-coding data so that the non-normality and non-constant variance of reported data can be corrected.

Treatment The controlled effect being assessed in an experiment for its influence on a measurable or observable outcome.

Type I error (False positive) Rejection of a true null hypothesis, P(Type I error) = significance level.

Type II error (False negative) Acceptance of a false null hypothesis, P(Type II error) = 1 – power/100.

Variability The level of variation present within collected data, also consistency and spread.

1

Introduction

1.1 INTRODUCTION

The origins of contemporary experimental design are generally attributed to Sir Ronald Fisher who published seminal work on statistical principles in the 1920s. The impact of his work was to become apparent in the late 1930s when terms such as **statistics**, **experimental design**, **treatment effect**, **randomisation**, **Analysis of Variance**, and **significance** and were to become recognised as synonymous with the efficient planning and analysis of data over a wide range of subject areas. In a famous address to the Indian Statistical Congress, Fisher was reported to have said:

> "To consult a statistician after an experiment is finished is often merely to ask him to conduct a *post mortem* examination. He can perhaps say what the experiment died of."

Fortunately, since then, there has been a prolific increase in the use of statistical methods and an appreciation of the merits of planned data collection and analysis. The methods have become the servant of the research and development community for the design and manufacture of new useful products. The methods have underpinned the discovery of fundamental knowledge and contributed to our understanding of the life sciences and social behaviour. Indeed, very few disciplines have made progress in the twentieth century and not been influenced by the principles of experimental design. In particular, experimental design has become the cornerstone of good statistical practice and internationally adopted by regulatory authorities and statutory bodies concerned with the safe and effective development of new processes and products. It should be realised that the term **experiment** is open to a very broad interpretation and covers any type of study, trial, or investigation where data are to be collected and assessed. It is not confined to the narrow interpretation of a laboratory experiment!

Despite these achievements the subject is still largely unexplored and under used. Many practitioners seldom get to implement the full range of rich techniques developed over the last 80 years by distinguished contributors such as Jerzy Neyman, Egon Pearson, Karl Pearson, Maurice Kendall, George Snedecor, John Tukey, Frank Yates, George Box, and Genichi Taguchi. This is either because of resource constraints or the benefits of the methods are not sufficiently promulgated. This comes at a time when data are increasingly generated and stored in large quantities using high speed computer systems and when there is a demand for the rapid assessment of complex multi-factor studies. Often the investigator will not be an expert in mathematics or statistics but someone from a completely different discipline with an interest in the application of experimental design to his or her problem domain.

Statistical practice covers a large number of statistical methods, and the key challenge to the modern day practitioner is to design the collection of data so that it