Design and Analysis of Experiments

Douglas C. Montgomery
Georgia Institute of Technology

Design and Analysis of Experiments

Douglas C. Montgomery
Georgia Institute of Technology

John Wiley & Sons
New York Santa Barbara London Sydney Toronto

Copyright © 1976, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

No part of this book may be reproduced by any means, nor transmitted, nor translated into a machine language without the written permission of the publisher.

Library of Congress Cataloging in Publication Data:

Montgomery, Douglas C
Design and analysis of experiments.

Bibliography: p.
Includes index.
1. Experimental design. I. Title.
QA279.M66 001.4'24 76-21075
ISBN 0-471-61421-1

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

This is an introductory textbook dealing with the statistical design and analysis of experiments. It is an outgrowth of lecture notes from a course in design of experiments that I have taught at the Georgia Institute of Technology over the past six years. The book is intended for readers who have completed a first course in statistical methods. The mathematical maturity that results from calculus would be helpful, but not essential, and some familiarity with matrix algebra is required in the latter part of Chapter 13 and Chapter 14.

Because of the relatively modest mathematical prerequisites, the book can be used in a second statistics course for undergraduate engineering, physical science, mathematics, biology, and social science students. The material can be covered more rapidly, with more emphasis on the mathematical aspects of the subject, in an experimental design course for first-year graduate students. There are numerical examples illustrating most of the design techniques, and this makes the book useful as a reference work for experimenters in various disciplines.

The book contains 15 chapters. Chapter 1 presents the basic philosophy of the statistical approach to experimental design. Chapter 2 reviews elementary statistical methods, and introduces terminology and notation used in subsequent chapters.

In Chapter 3, we begin the study of designs for experiments with a single factor. The analysis of variance is introduced as the appropriate method of statistical analysis. Chapters 4 and 5 continue the development of single-factor experiments, with randomized blocks, Latin squares, and related designs discussed in Chapter 4. Chapter 5 introduces incomplete block designs. Factorial designs are introduced in Chapter 6. Chapter 7 presents a set of rules for deriving computing formulas for sums of squares and expected mean squares for any balanced multifactor design.

The 2^k and 3^k factorial designs are introduced in Chapter 8. The 2^k and 3^k factorial designs may be run in incomplete blocks by sacrificing information on certain interactions. Procedures for constructing and analyzing these designs are given in Chapter 9. The relatively high cost of industrial experimentation has led to the extensive use of fractional 2^k and 3^k factorial designs, which are discussed in Chapter 10. The basic presentation of multifactor designs is continued in Chapter 11, which discusses nested arrangements. These are nonfactorial designs in which the levels of one factor are contained entirely within the levels of another factor. Chapter 12 illustrates how randomization restrictions are employed in multifactor experiments. An example of such a design would be running a factorial experiment in a randomized block.

Regression analysis is introduced in Chapter 13 as a methodology for the analysis of unplanned experiments. In Chapter 14 we discuss response surface methodology, a collection of mathematical and statistical techniques for determining the optimum operating conditions for industrial processes. This chapter concludes with a section on evolutionary operation, a process control method developed initially for chemical plants. The final chapter treats the analysis of covariance, which, like blocking, is a methodology for improving the precision of comparisons between treatments.

The book contains more material than can usually be covered comfortably in a first course. Hopefully, the instructor will be able to vary his course content from one offering to another, or perhaps discuss certain topics in greater depth, depending on class interest. There are problem sets at the end of each substantative chapter. These problems vary in scope from computational exercises designed to reinforce the fundamentals of the analysis of variance to extensions or elaborations of basic principles.

Many individuals have contributed toward the completion of the book. I thank Dr. R. N. Lehrer for his support and for his providing resources to develop the manuscript. I have benefited from reviews of the manuscript by Professors H. M. Wadsworth, K. S. Stephens, R. L. Rardin, R. G. Heikes, L. A. Johnson, and R. V. Fuller of the Georgia Institute of Technology; Professor A. L. Dorris of the University of Oklahoma; Professor J. J. Moder of the University of Miami; and Professor C. J. Tompkins of the University of Virginia. Their constructive criticism and suggestions have substantially improved the book. Various drafts of the manuscript have been used in my classes over the past two years, and I am grateful to the many students who corrected numerical and typographical errors, and provided critical evaluation. J. R. Black, G. N. Gollobin, R. E. Rosenthal, and R. D. Stewart were particularly helpful in this respect. I am indebted to Professor E. S. Pearson and the *Biometrika* Trustees, Wiley, Prentice-Hall, The Ronald Press, the editor of *Biometrics*, the editor of the *Annals of Statistics*, the editor of *Technometrics*, the Institute of Mathematical Statistics, and the American Statistical Association for permission to use copyrighted material. Finally, I thank Ms. Jeanie Hagen, Ms. Kaye Watkins, and Ms. Amelia L. N. Williams for typing the several drafts of the manuscript.

Douglas C. Montgomery
Atlanta, Georgia

试读结束：需要全本请在线购买： www.ertongbook.com

Contents

试读结束：需要全本请在线购买： www.ertongbook.com

Chapter One
Introduction

1-1 THE NEED FOR DESIGNED EXPERIMENTS

Experiments are carried out by investigators in all fields of study either to discover something about a particular process or to compare the effect of several conditions on some phenomena. For example, suppose a metallurgical engineer is interested in studying the effect of two different hardening processes, oil quenching and salt water quenching, on an aluminum alloy. Here the objective of the experimenter is to determine the quenching solution that produces the maximum hardness for this particular alloy. The engineer decides to subject a number of alloy specimens to each quenching medium and measure the hardness of the specimens after quenching. The average hardness of the specimens treated in each quenching solution will be used to determine which solution is best.

As we think about this experiment, a number of important questions come to mind.

1. Are these two solutions the only quenching media of potential interest?
2. Are there any other factors that might affect hardness that should be investigated or controlled in this experiment?
3. How many specimens of alloy should be tested in each quenching solution?

4. How should the specimens be assigned to the quenching solutions, and in what order should the data be collected?
5. What method of data analysis should be used?
6. What difference in average observed hardness between the two quenching media will be considered important?

All of these questions, and perhaps many others, will have to be satisfactorily answered before the experiment is performed.

In any experiment, the results and conclusions that can be drawn depend to a large extent on the manner in which the data were collected. To illustrate this point, suppose that the metallurgical engineer in the above experiment used specimens from one heat in the oil quench and specimens from a second heat in the salt water quench. Now when the mean hardness is compared, the engineer is unable to say how much of the observed difference is due to the quenching media and how much is due to inherent differences between the heats.[1] Thus the method of data collection has adversely affected the conclusions that can be drawn from the experiment.

1-2 BASIC PRINCIPLES OF EXPERIMENTAL DESIGN

If an experiment is to be performed most efficiently, then a scientific approach to planning the experiment must be considered. By the *statistical design of experiments*, we refer to the process of planning the experiment so that appropriate data will be collected, which may be analyzed by statistical methods resulting in valid and objective conclusions. The statistical approach to experimental design is necessary if we wish to draw meaningful conclusions from the data. When the problem involves data that are subject to experimental errors, statistical methodology is the only objective approach to analysis. Thus, there are two aspects to any experimental problem: the design of the experiment and the statistical analysis of the data. These two subjects are closely related, since the method of analysis depends directly on the design employed.

The two basic principles of experimental design are *replication* and *randomization*. By replication we mean a repetition of the basic experiment. In the metallurgical experiment above, a replication would consist of a specimen treated by oil quenching and a specimen tested by salt water quenching. Thus, if five specimens are treated in each quenching medium, we say that five *replicates* have been obtained. Replication has two important properties. First, it allows the experimenter to obtain an estimate of the experimental error. This estimate of error becomes a basic unit of measurement for determining whether observed differences in the data are really *statistically* different. Second, if the sample mean (e.g., $\bar{y}$) is used to estimate the effect of a factor in the experiment, then replication permits the experimenter to obtain a more precise es-

[1] A statistician would say that the effects of quenching media and heat were *confounded;* that is, the two effects cannot be separated.

timate of this effect; for if σ^2 is the variance of the data, and there are n replicates, then the variance of the sample mean is

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}$$

The practical implication of this is that if we had $n = 1$ replicate, and observed $y_1 = 145$ (oil quench) and $y_2 = 147$ (salt water quench) we would probably be unable to make satisfactory inferences about the effect of the quenching medium. That is, the observed difference could be due to experimental error. On the other hand, if n was reasonably large, and the experimental error was sufficiently small, then if we observed $\bar{y}_1 < \bar{y}_2$, we would be reasonably safe in concluding that salt water quenching produces a higher hardness in this particular aluminum alloy than does oil quenching.

Randomization is the cornerstone underlying the use of statistical methods in experimental design. By randomization we mean that both the allocation of the experimental material and the order in which the individual runs or trials of the experiment are to be performed are randomly determined. Statistical methods require that the observations (or errors) are independently distributed random variables. Randomization usually makes this assumption valid. By properly randomizing the experiment, we will also assist in "averaging out" the effects of extraneous factors that may be present. For example, suppose that the specimens in the above experiment are of slightly different thicknesses, and the effectiveness of the quenching medium may be affected by specimen thickness. If all the specimens subjected to the oil quench are thicker than those subjected to the salt water quench, then we may be continually handicapping one quenching medium over the other. Randomly assigning the specimens to the quenching media will alleviate this problem.

In order to use the statistical approach to designing and analyzing an experiment, it is necessary that everyone involved in the experiment have a clear idea in advance of exactly what is to be studied, how the data is to be collected, and at least a qualitative understanding of how this data is to be analyzed. An outline of the recommended procedure is as follows:

1. **Recognition of and statement of the problem.** This may seem to be a rather obvious point, but in practice it is often not simple to realize that a problem requiring experimentation exists, and to develop a clear and generally accepted statement of this problem. It is necessary to develop all ideas about the objectives of the experiment. A clear statement of the problem often contributes substantially to a better understanding of the phenomena and the final solution of the problem.

2. **Choice of factors and levels.** The experimenter must select the independent variables or factors to be investigated in the experiment. For example, in the hardness testing experiment described previously, the single factor is quenching media. The factors in an experiment may be either quantitative or qualitative. If they are quantitative, thought should be given as to how these factors are to be controlled at the desired values and measured. We must also select the values or levels of the factors to be used in the experiment. These levels may be chosen specifically, or selected at random from the set of all possible factor levels.

3. **Selection of a response variable.** In choosing a response or dependent variable, the experimenter must be certain that the response to be measured really provides information about the problem under study. Thought must also be given to how the response will be measured, and the probable accuracy of those measurements.

4. **Choice of experimental design.** This step is of primary importance in the experimental process. The experimenter must determine the difference in true response he wishes to detect and the magnitude of the risks he is willing to tolerate so that an appropriate sample size (number of replicates) may be chosen. He must also determine the order in which the data will be collected and the method of randomization to be employed. It is always necessary to maintain a balance between statistical accuracy and cost. Most recommended experimental designs are both statistically efficient and economical, so that the experimenter's efforts to obtain statistical accuracy usually result in economic efficiency. A mathematical model for the experiment must also be proposed, so that a statistical analysis of the data may be performed.

5. **Performing the experiment.** This is the actual data collection process. The experimenter should carefully monitor the progress of the experiment to insure that it is proceeding according to the plan. Particular attention should be paid to randomization, measurement accuracy, and maintaining as uniform an experimental environment as possible.

6. **Data analysis.** Statistical methods should be employed in analyzing the data from the experiment. Numerical accuracy is an important concern here, although present-day computers have largely relieved the experimenter from this problem, and simultaneously reduced the computational burden. Graphical methods are also frequently useful in the data analysis process.

7. **Conclusions and recommendations.** Once the data has been analyzed, the experimenter may draw conclusions or inferences about his results. The statistical inferences must be physically interpreted, and the practical significance of these findings evaluated. Then recommendations concerning these findings must be made. These recommendations may include a further round of experiments, as experimentation is usually an *iterative* process, with one experiment answering some questions and simultaneously posing others. In presenting his results and conclusions to others, the experimenter should be careful to minimize the use of unnecessary statistical terminology, and phrase his information as simply as possible. The use of charts and graphs is a very effective way to present important experimental results to management.

In this book we will concentrate primarily on step 4, the choice of experimental design, and step 6, the statistical analysis of the data. However, throughout the book we will emphasize the importance of the entire seven-step process.

1-3 HISTORICAL PERSPECTIVE

The late Sir Ronald A. Fisher was the innovator in the use of statistical methods in experimental design. For several years he was responsible for statistics and data analysis at the Rothamsted Agricultural Experiment Station in London, England. Fisher developed and first used the analysis of variance as the primary method of statistical analysis in experimental design. Frank Yates worked with Fisher at the Rothamsted station, and the two collaborated on many projects. Yates also became a primary contributor to the literature of experimental design. In 1933 Fisher took a professorship at the University of London. He later was on the faculty of Cambridge University, and held visiting professorships at several universities throughout the world.

Many of the early applications of experimental design methodology were in the agricultural and biological sciences. As a result, much of the terminology of the discipline is derived from this agricultural background. For example, an agricultural scientist may plant a variety of a crop in several plots, then apply different fertilizers or *treatments* to the plots, and observe the effect of the fertilizers on crop yield. Each plot will produce one observation on yield. An engineer or physical scientist, however, will think of the independent variable or *factor* that affects his response (rather than a treatment) and use the term *run* to characterize one observation. However, much experimental design terminology, such as "treatment," "plot," and "block" have lost their strictly agricultural connotation, and are in wide use in many fields of application. We will use the phrases "levels of the factor" and "treatment" interchangeably. Sometimes "treatment combination" will be used to denote a particular combination of factor levels to be used in one run of the experiment.

Modern-day experimental design methods are widely employed in all fields of inquiry. Agricultural science, biology, medicine, the engineering sciences, the physical sciences, and the social sciences are disciplines where the statistical approach to the design and analysis of experiments is an accepted practice.

1-4 AN EXAMPLE OF A DESIGNED EXPERIMENT

To illustrate some of the preceding ideas, we present an example of a designed experiment. A hardness testing machine presses a rod with a pointed tip into a metal specimen with a known force. By measuring the depth of the depression caused by the tip, the hardness of the specimen is determined. Two different tips are available for this machine, and although the precision (variability) of the two tips seems to be the same, it is suspected that one tip produces different hardness readings than the other.

An experiment could be performed as follows. A number of metal specimens (e.g., 10) could be randomly selected. Half of these specimens could be tested by tip 1 and the other half by tip 2. The exact assignment of specimens to tips would be randomly determined. After the hardness data have been collected, the average hardness of the two samples could be compared using the t test. That is, if n_1 specimens are tested with tip 1 and n_2 specimens are tested with tip 2, and $\bar{y}_1$ and $\bar{y}_2$ are the resulting average hardness from tips 1 and 2, respectively, then assuming that the hardness data is normally distributed, the test statistic is

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{1-1}$$

where

$$S_p = \left[\frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} \right]^{1/2} \quad (1\text{-}2)$$

is the pooled or combined estimate of the variability (experimental error). Formally stated, the hypothesis we are testing is

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

where μ_1 and μ_2 are the true means of the distributions of hardness values produced by tips 1 and 2, respectively. We will reject the null hypothesis $H_0: \mu_1 = \mu_2$ if

$$|t_0| > t_{\alpha/2, n_1+n_2-2}$$

where $t_{\alpha/2, n_1+n_2-2}$ is the upper $\alpha/2$ percentage point of the t distribution with $n_1 + n_2 - 2$ degrees of freedom.

A little reflection will reveal a serious disadvantage of this design. Suppose the metal specimens were cut from different bar stock that were produced in different heats, or were not exactly homogeneous in some other way that might affect the hardness. This lack of homogeneity between specimens will contribute to the variability of the hardness measurements and tend to inflate the experimental error, thus making a true difference between tips harder to detect.

To protect against this possibility, consider an alternate experimental design. Assume that each specimen is large enough so that *two* hardness determinations may be made on it. This alternative design would consist of dividing each specimen into two parts, then randomly assigning one tip to one half of each specimen and the other tip to the remaining half. The order in which the tips are tested for a particular specimen would also be randomly selected. The experiment, when performed according to this design with 10 specimens, produced the data shown in Table 1-1.

Table 1-1 Data for the Hardness Testing Experiment

Specimen	Tip 1	Tip 2
1	7	6
2	3	3
3	3	5
4	4	3
5	8	8
6	3	2
7	2	4
8	9	9
9	5	4
10	4	5

试读结束：需要全本请在线购买： www.ertongbook.com