

Statistics for Experimenters

*An Introduction to Design,
Data Analysis,
and Model Building*

GEORGE E. P. BOX

WILLIAM G. HUNTER

J. STUART HUNTER

Statistics for Experimenters

*An Introduction to Design,
Data Analysis,
and Model Building*

GEORGE E. P. BOX

WILLIAM G. HUNTER

J. STUART HUNTER

John Wiley & Sons
New York • Chichester • Brisbane • Toronto

Copyright © 1978 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Sections 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data

Box, George E. P.

Statistics for experimenters.

(Wiley series in probability and mathematical statistics)

Includes index.

1. Experimental design. 2. Analysis of variance. I. Hunter, William Gordon, 1937– joint author. II. Hunter, J. Stuart, 1923– joint author. III. Title.

QA279.B68 001.4'24 77-15087
ISBN 0-471-09315-7

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

Collaboration with research workers has been a source of satisfaction to us as practicing statisticians. It has put us in touch with a wide variety of investigations and has stimulated us by intellectual contact with many investigators. We have certainly learned from them, and they have sometimes learned from us. This book owes much to such interaction. It is written for those who collect data and try to make sense of it, and it is an introduction to those ideas and techniques that we have found especially useful.

For a number of years we have used preliminary versions (in the form of notes) of this book in teaching engineers, chemists, biologists, statisticians, and other scientists at the University of Wisconsin and Princeton University. We have also used this material in courses for professional societies, industry, and government. It is neither a cookbook nor a textbook on mathematical statistics. It is an introduction to the philosophy of experimentation and the part that statistics plays in experimentation.

Actual problems are never so straightforward that they can safely be solved mechanically. Therefore we emphasize the necessity for thinking about the real nature of the scientific problem itself, for mulling over data plots and other graphical displays, as well as for understanding potentially useful statistical principles and their practical consequences. We point out possible difficulties caused by the violation of assumptions, such as the lack of independence of data. Furthermore we discuss ways by which such difficulties may be overcome. To illustrate principles we use both real and constructed data. For example, when the analysis of variance is first introduced, we simplify a real set of data so that all the values required in the analysis are whole numbers. Special emphasis is placed on the design of experiments because this is the most valuable aspect of statistical method. Frequently conclusions are easily drawn from a well-designed experiment, even when rather elementary methods of analysis are employed. Conversely, even the most sophisticated statistical analysis cannot salvage a badly designed experiment.

Statistical theory is introduced as it becomes necessary. Readers are assumed to have no previous knowledge of the subject; and although the book presents a properly sophisticated view of the philosophy of scientific investigation and the part played by statistical design and analysis, the mathematics needed is elementary. In particular, calculus does not appear in the main body of the text and is not needed to understand and use the book. Even more important than learning about statistical techniques is the development of what might be called a capability for *statistical thinking*. We hope this book will serve as a vehicle for this purpose. To help convey this facet of statistics, which is a challenge to the instructor, we have included several different types of questions and problems. In addition to exercises throughout the text, there are review questions at the end of each chapter, and problems at the end of each part of the book.

In addition to experimenters, others will find this book useful, for example, those who assess reported findings (Do the data really support these claims?), managers who direct research projects (What is the best way to approach this investigation?), and statisticians who work with experimenters (What is the best method of design and analysis for this problem?).

A typical one-semester college course might include fairly complete coverage of the first three parts of the book (Chapters 1–13) and a selection of topics from the fourth part (Chapters 14–18). Although this book is primarily intended for use as a text, we have also kept in mind the reader who will use it for self-instruction and/or as a reference.

Special Notes to the Reader

The questions at the end of each chapter reflect its contents and can be used in two ways: You can consider them for review *after* reading the chapter, or you can study them *before* reading the chapter to help identify key points. You may also find it helpful to use the problems listed at the end of each part of the book to guide your reading.

As you apply the ideas in this book, and especially if you meet with spectacular success or failure, we shall be interested to learn of your experience. We have tried to write a book that is useful and clear. If you have suggestions about how it can be improved, please write to us. As Chapter 1 suggests, we believe in iteration.

Acknowledgments

The manuscript was begun at Princeton University and continued at the University of Wisconsin, Madison, at both the Department of Statistics and the Mathematics Research Center. We record our thanks to Bovas Abraham,

Steve Bailey, Gina Chen, Johannes Ledolter, Greta Ljung, and Lars Pallesen for helping with the calculations and for checking the final manuscript. We are also indebted to the hundreds of students who used the notes on which this book is based and their proposed improvements.

We are grateful to Professor E. S. Pearson and the Biometrika Trustees for permission to reprint condensed and adapted forms of various tables from *Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., 1966. These are listed at the end of this book as Tables A, B1, C, D, F, and G.

We are grateful also to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Longman Group Ltd., London, for permission to partially reprint Table III from their book *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed., 1974.

Our thanks are due to William G. Cochran and Gertrude M. Cox for allowing us to reproduce material on incomplete blocks designs from their classic book, *Experimental Designs*.

Finally, we are glad to acknowledge our debt to Mary Esser, Mary Arthur, and Doris A. Whitmore, who carefully typed the final manuscript.

G. E. P. BOX
W. G. HUNTER
J. S. HUNTER

Madison, Wisconsin
Princeton, New Jersey
March 1978

Greek Alphabet

A α	alpha	N ν	nu
B β	beta	Ξ ξ	xi
Γ γ	gamma	Ο ο	omicron
Δ δ	delta	Π π	pi
Ε ε	epsilon	Ρ ρ	rho
Ζ ζ	zeta	Σ σ	sigma
Η η	eta	Τ τ	tau
Θ θ	theta	Υ υ	upsilon
Ι ι	iota	Φ φ	phi
Κ κ	kappa	Χ χ	chi
Λ λ	lambda	Ψ ψ	psi
Μ μ	mu	Ω ω	omega

Contents

1	SCIENCE AND STATISTICS	1
1.1	The Learning Process	1
1.2	The Role of Experimental Design	4
1.3	Difficulties Mitigated by Statistical Methods	7
1.4	A Typical Investigation	9
1.5	How to Use Statistical Techniques	14
PART I COMPARING TWO TREATMENTS		
2	USE OF EXTERNAL REFERENCE DISTRIBUTION TO COMPARE TWO MEANS	21
2.1	Relevant Reference Sets and Distributions	21
2.2	Theory: Probability Distributions, Parameters, and Statistics	24
2.3	The Industrial Experiment: External Reference Distribution	31
2.4	Theory: Normal and t Distributions	38
2.5	The Industrial Experiment: An External Reference Distribution Based on the t Distribution	51
	Appendix 2A Calculation of the Sample Average, Sample Variance, and Sample Standard Deviation by Coding Data	53
3	RANDOM SAMPLING AND THE DECLARATION OF INDEPENDENCE	57
3.1	Theory: Statistical Dependence and Independence and the Random Sampling Model	57
		xi

3.2	The Industrial Experiment: Reference Distribution Based on Random Sampling Model, External Value for σ	74
3.3	The Industrial Experiment: Reference Distribution Based on Random Sampling Model, Internal Estimate of σ	76
3.4	Summary: What Have We Learned from the Industrial Experiment Example?	82
Appendix 3A	Mean and Variance of a Linear Combination of Observations	87
Appendix 3B	Robustness of Some Statistical Procedures	89
Appendix 3C	Fisher's Concept of Sufficiency	91
4	RANDOMIZATION AND BLOCKING WITH PAIRED COMPARISONS	93
4.1	Randomization to the Rescue: Tomato Plant Example	93
4.2	Randomized Paired Comparison Design: Boys' Shoes Example	97
4.3	Blocking and Randomization	102
4.4	Noise Structure, Models, and Randomization	104
4.5	Summary: Comparison, Replication, Randomization, and Blocking in Simple Comparative Experiments	105
5	SIGNIFICANCE TESTS AND CONFIDENCE INTERVALS FOR MEANS, VARIANCES, PROPORTIONS, AND FREQUENCIES	107
5.1	A More Detailed Discussion of Significance Tests	107
5.2	Confidence Intervals for a Difference in Means: Paired Comparison Design	110
5.3	Confidence Intervals for a Difference in Means: Unpaired Design	115
5.4	Inferences about Variances of Normally Distributed Data	117
5.5	Inferences about Proportions: The Binomial Distribution	123
5.6	Inferences about Frequencies: The Poisson Distribution	137
5.7	Contingency Tables and Tests of Association	145
	PROBLEMS FOR PART I	152

PART II COMPARING MORE THAN TWO TREATMENTS

6	EXPERIMENTS TO COMPARE k TREATMENT MEANS	165
6.1	Blood Coagulation Times with Four Different Diets	165
6.2	Estimating the Amount of Variation Within and Between Treatments	167
6.3	The Arithmetic and Geometry of the Analysis of Variance Table	170
6.4	Decomposition of the Observations Implied by the Analysis	175
6.5	Diagnostic Checking of the Basic Model	182
6.6	Use of the Analysis of Variance Table	187
6.7	Use of a Reference Distribution to Compare Means	190
6.8	Summary	193
	Appendix 6A Shortcut Method for Constructing the Analysis of Variance Table	194
	Appendix 6B Vectors and Geometry Associated with the Analysis of a Sample	197
	Appendix 6C Multiple Comparisons	203
7	RANDOMIZED BLOCKS AND TWO-WAY FACTORIAL DESIGNS	208
7.1	Example: Comparison of Four Variants of a Penicillin Production Process	209
7.2	A Model with Corresponding Decomposition of Observations	210
7.3	Implications of the Additive Model	218
7.4	Diagnostic Checking of the Model	220
7.5	Use of the Analysis of Variance Table	223
7.6	The Use of Reference Distributions To Compare Individual Means	226
7.7	A Two-Way (Factorial) Design	228
7.8	Simplification and Increased Sensitivity from Transformation	231
7.9	Likelihood Estimation of the Transformation	239
7.10	Summary	241
	Appendix 7A Calculations for Constructing Analysis of Variance Table for Randomized Block Design	241

Appendix 7B	Algebraic Demonstration of the Additivity of the Sums of Squares in a Randomized Block	243
8	DESIGNS WITH MORE THAN ONE BLOCKING VARIABLE	245
8.1	Latin Square Designs: Automobile Emissions and Synthetic Yarn Examples	245
8.2	Graeco- and Hyper-Graeco-Latin Squares: First Wear Testing Example	255
8.3	Balanced Incomplete Block Designs: Second Wear Testing Example	258
Appendix 8A	Some Useful Latin Squares and How to Use Them to Construct Graeco-Latin and Hyper-Graeco-Latin Square Design	261
Appendix 8B	Analysis of Variance for $k \times k$ Latin Square Designs with r Replicates	263
Appendix 8C	Some Useful Balanced Incomplete Block Designs	269
Appendix 8D	Analysis of Variance and Computation of Adjusted Treatment Averages for Balanced Incomplete Block Designs	275
	PROBLEMS FOR PART II	281
	PART III MEASURING THE EFFECTS OF VARIABLES	
9	EMPIRICAL MODELING	291
9.1	Mathematical Models	291
9.2	Geometric Representation of Empirical Relationships	296
9.3	The Problem of Experimental Design	298
9.4	Comprehensive Versus Sequential Approach to Experimental Investigations	303
10	FACTORIAL DESIGNS AT TWO LEVELS	306
10.1	General Factorial Designs and Designs at Two Levels	306
10.2	An Example of a 2^3 Factorial Design: Pilot Plant Investigation	307

CONTENTS	xv
10.3 Calculation of Main Effects	309
10.4 Interaction Effects	313
10.5 Interpretation of Results	317
10.6 Calculation of Standard Errors for Effects Using Replicated Runs	319
10.7 Quicker Methods for Calculating Effects	322
10.8 A 2^4 Factorial Design: Process Development Study	324
10.9 Analysis of Factorials Using Normal Probability Paper	329
10.10 Transformation of Data from Factorial Designs	334
10.11 Blocking	336
10.12 Summary	342
Appendix 10A Yates's Algorithm	342
Appendix 10B More on Blocking Factorial Designs	344
11 MORE APPLICATIONS OF FACTORIAL DESIGNS	352
11.1 Example 1: The Effects of Three Variables on Clarity of Film	352
11.2 Example 2: The Effects of Three Variables on Physical Properties of a Polymer Solution	353
11.3 Example 3: Development of Screening Facility for Storm Water Overflows	354
11.4 Example 4: Simple Factorials Used Sequentially in Evolutionary Operation—Petrochemical Plant	362
11.5 Example 5: Simple Factorials Used Sequentially in Evolutionary Operation—Polymer Unit	365
11.6 Summary	368
Appendix 11A A Suggested Exercise	368
12 FRACTIONAL FACTORIAL DESIGNS AT TWO LEVELS	374
12.1 Redundancy	374
12.2 A Half-Fraction of a 2^5 Design: Reactor Example	376
12.3 Construction and Analysis of Half-Fractions: Reactor Example	381
12.4 The Concept of Design Resolution: Reactor Example	385
12.5 Resolution III Designs: Bicycle Example	390
12.6 Resolution IV Designs: Injection Molding Example	398
12.7 Elimination of Block Effects in Fractional Designs	404

12.8	Designs of Resolution V and Higher	407
12.9	Summary	409
Appendix 12A	Structure of the Fractional Designs	409
Appendix 12B	Choosing Additional Runs To Resolve Ambiguities from Fractional Factorials	413
13	MORE APPLICATIONS OF FRACTIONAL FACTORIAL DESIGNS	419
13.1	Example 1: Effects of Five Variables on Some Properties of Cast Films	419
13.2	Example 2: Stability of New Product	422
13.3	Example 3: Bottleneck at the Filtration Stage of an Industrial Plant	424
13.4	Example 4: Sensitivity Analysis of a Simulation Model—Controller–Aircraft System	429
13.5	Summary	432
	PROBLEMS FOR PART III	434
	PART IV BUILDING MODELS AND USING THEM	
14	SIMPLE MODELING WITH LEAST SQUARES (REGRESSION ANALYSIS)	453
14.1	One-Parameter Model (Straight Line through the Origin): Aerosol Example	453
14.2	Two-Parameter Model: Impurity Example	462
14.3	Straight Line Model: Welding Example	473
14.4	General Case for Models Linear in the Parameters	479
14.5	Polynomial Model: Growth Rate Example	480
14.6	Nonlinear Model: Biochemical Oxygen Demand Example	483
14.7	Hazards of Fitting Regression Equations to Happenstance Data	487
Appendix 14A	Why Do the Normal Equations Yield Least Squares Estimates?	498
Appendix 14B	Matrix Version of the Normal Equations	501

Appendix 14C	Analysis of Factorials, Botched and Otherwise	503
Appendix 14D	Unweighted and Weighted Least Squares	505
15	RESPONSE SURFACE METHODS	510
15.1	Weakness of Classical One-Variable-at-a-Time Strategy: Chemical Example	510
15.2	Illustration of Response Surface Methodology: Chemical Example	513
15.3	A Specification Problem	526
15.4	Maxima, Ridges, and Canonical Analysis	526
15.5	Applications of Response Surface Methods	534
15.6	Summary	535
16	MECHANISTIC MODEL BUILDING	540
16.1	Empirical and Mechanistic Models	540
16.2	Possible Advantages of Mechanistic Models	544
16.3	Techniques for Mechanistic Modeling	546
16.4	The Model-Building Process	548
16.5	Model Testing with Diagnostic Parameters	550
16.6	Importance of Plotting Data in the Age of Computers	552
16.7	Summary	552
17	STUDY OF VARIATION	556
17.1	Graphs and Control Charts: Impurity Determination Example	556
17.2	Transmission of Error	563
17.3	Variance Components: Pigment Paste Example	571
Appendix 17A	Calculating Variance Components from an Analysis of Variance Table	581
18	MODELING DEPENDENCE: TIME SERIES	584
18.1	The Industrial Data of Chapter 2 Reconsidered as a Time Series	585
18.2	Statistical Modeling Revisited	588

18.3	Forecasting: Refrigerator Sales Example	591
18.4	Feedback Control: Dye Level Example	598
18.5	Intervention Analysis: Los Angeles Air Pollution Example	602
Appendix 18A	Derivations of Equation 18.4	604

PROBLEMS FOR PART IV

APPENDIX: TABLES

INDEX

606	Weakness of Classical One-Variable at a Time
629	Chemical Example
645	Illustration of Response Surface Methodology: Chemical Example
	A Specification Problem
	Maxima, Ridges, and Canonical Analysis
	Applications of Response Surface Methods
	Summary
	MECHANISTIC MODEL BUILDING
	16.1 Empirical and Mechanistic Models
	16.2 Possible Advantages of Mechanistic Models
	16.3 Techniques for Mechanistic Modeling
	16.4 The Model-Building Process
	16.5 Model Testing with Diagnostic Parameters
	16.6 Importance of Plotting Data in the Age of Computers
	16.7 Summary
	STUDY OF VARIATION
	17.1 Graphs and Control Charts: Impurity Determination Example
	17.2 Transmission of Error
	17.3 Variance Components: Pigment Paste Example
	Appendix 17A Calculating Variance Components from an Analysis of Variance Table
	MODELING DEPENDENCE: TIME SERIES
	18.1 The Industrial Data of Chapter 2 Reconsidered as a Time Series
	18.2 Statistical Modeling Revisited

CHAPTER 1

Science and Statistics

1.1. THE LEARNING PROCESS

Scientific research is a process of guided learning. The object of statistical methods is to make that process as efficient as possible.

Learning is advanced by the iteration illustrated in Figure 1.1. An initial hypothesis leads by a process of *deduction* to certain necessary consequences that may be compared with data. When consequences and data fail to agree, the discrepancy can lead, by a process called *induction*, to modification of the hypothesis. A second cycle in the iteration is thus initiated. The consequences of the modified hypothesis are worked out and again compared with data (old or newly acquired*) that in turn can lead to further modification and gain of knowledge.

This process of learning can also be depicted as a feedback loop (Figure 1.2) in which the discrepancy between the data and the consequences of hypothesis H_1 leads to the modified hypothesis H_2 , H_2 leads to H_3 , and so on.

Nursery School Example

In a certain nursery school the teacher was pouring juice into cups for the children, who were sitting around a table. The teacher asked what would happen if, rather than stopping, she just kept pouring the juice into one of the cups. One child said the juice would fly *up* to the sky, and all the other children, except one, agreed. The one dissenting child (very likely drawing on her own past experience) said the juice would overflow and run *down* onto the table. The teacher took the opportunity to perform the experiment and demonstrate what actually happens. Learning took place that morning.

* In most of our applications the data acquiring process is scientific experimentation, but it could be a walk to the library to unearth already existing information or the conduct of a sample survey.

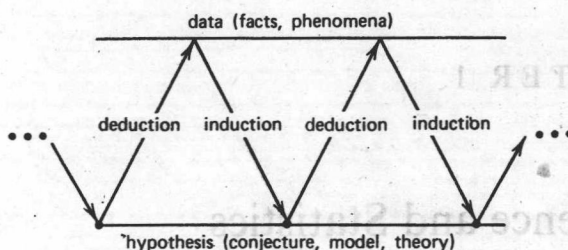


FIGURE 1.1. The iterative learning process.

Chemical Example

A chemist might have the following learning experience.

Hypothesis 1 Because of certain properties of a newly discovered catalyst, its presence in a particular reaction mixture would deduced probably cause chemical *A* to combine with chemical *B* to form, in high yield, a valuable product *C*.
its consequences

The chemist has a tentative hypothesis and deduces its consequences, but he has no data on which to verify or deny its truth because, as far as he can tell from discussion with colleagues and careful examination of the literature, no one has ever performed the operation in question. He therefore decides to run some experiments.

Experimental design 1 A run is made at carefully selected reaction conditions. In particular, the chosen reaction temperature is 600°C.

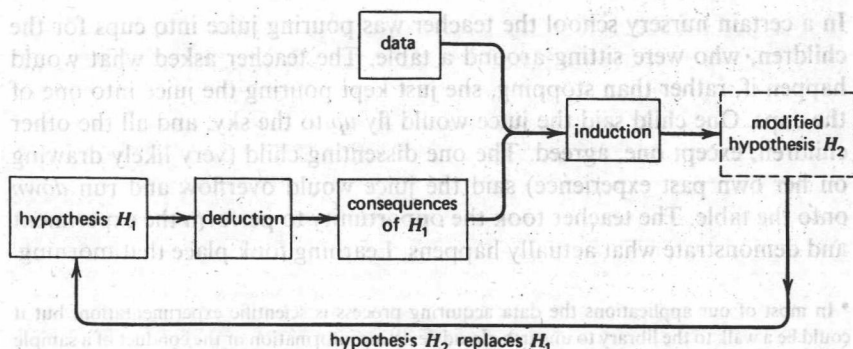


FIGURE 1.2. The learning process as a feedback loop.