APPLICATIONS OF STATISTICS TO INDUSTRIAL EXPERIMENTATION

CUTHBERT DANIEL

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

This Page Intentionally Left Blank

APPLICATIONS OF STATISTICS TO INDUSTRIAL EXPERIMENTATION

CUTHBERT DANIEL

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto • Singapore

This Page Intentionally Left Blank

Applications of Statistics to Industrial Experimentation

This Page Intentionally Left Blank

A NOTE TO THE READER
This book has been electronically reproduced from digital information stored at John Wiley & Sons, Inc. We are pleased that the use of this new technology will enable us to keep works of enduring scholarly value in print as long as there is a reasonable demand for them. The content of this book is identical to previous printings.

Published by John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data:

Daniel, Cuthbert.
Applications of statistics to industrial experimentation.

Bibliography: p. 289
Includes indexes.
1. Experimental design.
2. Research, Industrial—Statistical methods. I. Title.

T57.37.D36 607'.2 76-2012
ISBN 0-471-19469-7

10 9 8 7

To Janet

This Page Intentionally Left Blank

Preface

An experiment is an intervention in the operation of a working system. It is always done to learn about the effect of the change in conditions. Some conditions must be controlled; some, at least one, must be varied deliberately, not just passively observed. To avoid Chantecler's mistake, the variation should not be regular. (You will remember that Rostand's cock thought it was his crowing that made the sun rise.) All industrial experiments are interventions; unfortunately not all are irregularly timed interventions.

It is impossible to make any very general statistical statements about industrial experiments. No claim is made here for the universal applicability of statistical methods to the planning of such experiments. Rather, we proceed by examples and by modest projections to make some judgments on some sorts of industrial experiments that may gain from statistical experience.

Industrial experiments may be classified in several ways that carry implications for statistical thinking. First, I put J. W. Tukey's distinction between confirmation and exploration experiments, which might well be extended by the small but important classification of fundamental, or creative, or stroke-of-genius experiments. This book deals almost entirely with confirmatory experiments, a little with exploratory ones, and not at all with the last type. Confirmation experiments are nearly always done on a working system and are meant to verify or extend knowledge about the response of the system to varying levels or versions of the conditions of operation. The results found are usually reported as point- or confidence-interval statements, not as significance tests or P-values.

A second way of classifying experiments is based on the distance of their objectives from the market. As we get closer to being ready to go into production (or to making a real change in production operations), it becomes more important to have broadly based conclusions, covering the effects of realistic ranges of inputs, operating conditions, on all properties of the product. The farther we are to the right on the God–Mammon scale, the more useful large-scale multifactor experiments are likely to be.

A third classification involves continuity of factors. If most factors in an experimental situation are continuously variable and are controllable at predetermined levels, the whole range of response surface methodology becomes available. These procedures are only cursorily discussed here, since there are already many excellent expositions in print. When many factors are orderable in their levels, but not measurable, the response surface methods become less useful. When many factors are discrete-leveled and unorderable, one's thinking and one's designs necessarily change to accommodate these facts. It is with these latter types of situations that this work is mainly concerned.

A fourth classification distinguishes between experimental situations in which data are produced sequentially and those in which many results are produced simultaneously, perhaps after a lapse of time. Pilot plants, full-scale factory operations, and even bench work on prototype equipment usually produce one result at a time. Storage tests, and clinical trials on slowly maturing diseases are examples of situations that are intrinsically many at a time, not one at a time. They are always multiple simultaneous trials since a long time may be needed to fill in omissions. A very large number of such experiments have been carried out, and dozens have been published. They are strongly isomorphic with the corresponding agricultural factorial experiments. At the one-at-a-time end of this scale I believe but cannot prove that some statistical contribution is to be expected. No examples of completed sets can be given.

Experiments vary in their sensitivity. In some situations the effect of interest Δ is four or more times the error standard deviation σ of the system, so that $\Delta/\sigma \geqslant 4$. In such cases, small numbers of trials (runs, tests, subexperiments) are required, and replication is supererogatory. This happens most commonly in physical sciences, and in bench work when the experimental setup is familiar and stable. At the other extreme are situations in which $\Delta/\sigma \leqslant 1$, as is common in the biological sciences, including clinical trials, and in work on large-scale, even plant-wide, experiments, where uncontrollable variation is always present and small improvements are commercially important. Statistical methods can be well adjusted to this whole gamut, and the details of this coverage will be given in several chapters.

The book should be of use to experimenters who have some knowledge of elementary statistics and to statisticians who want simple explanations, detailed examples, and a documentation of the variety of outcomes that may be encountered.

CUTHBERT DANIEL

Rhinebeck, New York
March 1976

Acknowledgments

Parts of this work have been discussed and criticized in statistical colloquia at the universities of Princeton, Yale, and Connecticut at Storrs and at the Courant Institute of New York University. Harry Smith and J. S. Hunter have given me valuable detailed criticism of an earlier draft. Also, F. Anscombe, W. G. Cochran, O. Kempthorne, B. H. Margolin, R. W. Kennard, F. Mosteller, H. Scheffé, and H. F. Smith have made searching comments. I am grateful for all of this help and have tried to incorporate it into a better book.

My clients over the past 30 years have provided the ground, the basis, the occasion for all that I may have learned in applied statistics. Without exception they have been patient, intelligent, and generous. I think especially of the research and development staffs of Kellex–Union Carbide Chemicals Corporation at Oak Ridge, of Procter and Gamble, United States Steel Corporation, M. W. Kellogg, General Foods, American Oil Company–Panamerican Petroleum, Itek, Okonite, Interchemical Corporation, Consumers Union, and Technicon Instruments Corporation.

I am deeply indebted also to Ms. B. Shube of Wiley-Interscience and to many others on the staff of that company for their patience and care in editing and seeing this work through the press. It is a pleasure to thank the printers and compositors who have so often seen how to make a clear table from an unclear manuscript. In addition, I want to thank the publisher's reviewers, who gave me many good suggestions.

S. Addelman and the editors of *Technometrics* have given me permission to reprint several tables and one full article (all identified in the text).

Finally, H. Scheffé has provided, over the past twenty-five years, aid and encouragement beyond praise and almost beyond tally. I find that I have four thick folders in my files, all labeled "Scheffé, Current."

C. D.

This Page Intentionally Left Blank

Contents

Applications of Statistics
to Industrial Experimentation

WILEY SERIES IN PROBABILITY
AND MATHEMATICAL STATISTICS

Probability and Mathematical Statistics

ADLER • The Geometry of Random Fields
ANDERSON • The Statistical Analysis of Times Series
ANDERSON • An Introduction to Multivariate Statistical Analysis
ARAUJO and GINE • The Central Limit Theorem for Real and Banach Valued Random Variables
ARNOLD • The Theory of Linear Models and Multivariate Analysis
BARLOW, BARTHOLOMEW, BREMNER, and BRUNK • Statistical Inference Under Order Restrictions
BARNETT • Comparative Statistical Inference
BHATTACHARYYA and JOHNSON • Statistical Concepts and Methods
BILLINGSLEY • Probability and Measure
CASSEL, SARNDAL, and WRETMAN • Foundations of Inference in Survey Sampling
DE FINETTI • Theory of Probability, Volumes I and II
DOOB • Stochastic Processes
FELLER • An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition*, Revised
FELLER • An Introduction to Probability Theory and Its Applications, Volume II, *Second Edition*
FULLER • Introduction to Statistical Time Series
GRENANDER • Abstract Inference
HANNAN • Multiple Time Series
HANSEN, HURWITZ, and MADOW • Sample Survey Methods and Theory, Volumes I and II
HARDING and KENDALL • Stochastic Geometry
HOEL • Introduction to Mathematical Statistics, *Fourth Edition*
HUBER • Robust Statistics
IOSIFESCU • Finite Markov Processes and Applications
ISAACSON and MADSEN • Markov Chains
KAGAN, LINNIK, and RAO • Characterization Problems in Mathematical Statistics
KENDALL and HARDING • Stochastic Analysis
LAHA and ROHATGI • Probability Theory
LARSON • Introduction to Probability Theory and Statistical Inference, *Second Edition*
LARSON • Introduction to the Theory of Statistics
LEHMANN • Testing Statistical Hypotheses
MATHERON • Random Sets and Integral Geometry
MATTHES, KERSTAN, and MECKE • Infinitely Divisible Point Processes
PARZEN • Modern Probability Theory and Its Applications
PURI and SEN • Nonparametric Methods in Multivariate Analysis
RAGHAVARAO • Constructions and Combinatorial Problems in Design of Experiments
RANDLES and WOLFE • Introduction to the Theory of Nonparametric Statistics
RAO • Linear Statistical Inference and Its Applications, *Second Edition*
ROHATGI • An Introduction to Probability Theory and Mathematical Statistics
SCHEFFE • The Analysis of Variance
SEBER • Linear Regression Analysis
SERFLING • Approximation Theorems of Mathematical Statistics
TJUR • Probability Based on Radon Measures

Probability and Mathematical Statistics (Continued)

WILLIAMS • Diffusions, Markov Processes, and Martingales, Volume I: Foundations

ZACKS • Theory of Statistical Inference

Applied Probability and Statistics

ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG • Statistical Methods for Comparative Studies

ARTHANARI and DODGE • Mathematical Programming in Statistics

BAILEY • The Elements of Stochastic Processes with Applications to the Natural Sciences

BAILEY • Mathematics, Statistics and Systems for Health

BARNETT and LEWIS • Outliers in Statistical Data

BARTHOLEMEW • Stochastic Models for Social Processes, *Second Edition*

BARTHOLOMEW and FORBES • Statistical Techniques for Manpower Planning

BECK and ARNOLD • Parameter Estimation in Engineering and Science

BELSLEY, KUH, and WELSCH • Regression Diagnostics: Identifying Influential Data and Sources of Collinearity

BENNETT and FRANKLIN • Statistical Analysis in Chemistry and the Chemical Industry

BHAT • Elements of Applied Stochastic Processes

BLOOMFIELD • Fourier Analysis of Time Series: An Introduction

BOX • R. A. Fisher, The Life of a Scientist

BOX and DRAPER • Evolutionary Operation: A Statistical Method for Process Improvement

BOX, HUNTER, and HUNTER • Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building

BROWN and HOLLANDER • Statistics: A Biomedical Introduction

BROWNLEE • Statistical Theory and Methodology in Science and Engineering, *Second Edition*

BURY • Statistical Models in Applied Science

CHAMBERS • Computational Methods for Data Analysis

CHATTERJEE and PRICE • Regression Analysis by Example

CHERNOFF and MOSES • Elementary Decision Theory

CHOW • Analysis and Control of Dynamic Economic Systems

CLELLAND, BROWN, and deCANI • Basic Statistics with Business Applications, *Second Edition*

COCHRAN • Sampling Techniques, *Third Edition*

COCHRAN and COX • Experimental Designs, *Second Edition*

CONOVER • Practical Nonparametric Statistics, *Second Edition*

CORNELL • Experiments with Mixtures: Designs, Models and The Analysis of Mixture Data

COX • Planning of Experiments

DANIEL • Biostatistics: A Foundation for Analysis in the Health Sciences, *Second Edition*

DANIEL • Applications of Statistics to Industrial Experimentation

DANIEL and WOOD • Fitting Equations to Data: Computer Analysis of Multifactor Data, *Second Edition*

DAVID • Order Statistics, *Second Edition*

DEMING • Sample Design in Business Research

DODGE and ROMIG • Sampling Inspection Tables, *Second Edition*

DRAPER and SMITH • Applied Regression Analysis

DUNN • Basic Statistics: A Primer for the Biomedical Sciences, *Second Edition*

DUNN and CLARK • Applied Statistics: Analysis of Variance and Regression

ELANDT-JOHNSON • Probability Models and Statistical Methods in Genetics

ELANDT-JOHNSON and JOHNSON • Survival Models and Data Analysis

continued on back

This Page Intentionally Left Blank

CHAPTER 1

Introduction

1.1. THE RANGE OF INDUSTRIAL RESEARCH

The connections between scientific research and industrial research are sometimes very close. In studying a new industrial use for the water-gas shift reaction, for example, industrial research workers would depend heavily on the theoretical and experimental results in the technical literature. In producing a new modification of a familiar dyestuff with somewhat improved lightfastness, one industrial organic chemist would start with a careful theoretical study and search for the relevant literature. Another equally able chemist might prefer a wider search of alternatives directly in the laboratory. In attempting to find acceptable operating conditions to make a new petrochemical, it might well be discovered that *no* basis for a theory exists until a considerable volume of laboratory work has been completed.

A wide spectrum of degrees of empiricism already exists, then, in industrial research. The word *theory* is used with entirely different references in different parts of this spectrum. The word may be almost a term of derogation when used by a chemist working on a problem requiring a high degree of empiricism, to describe the work of another who has a good mathematical background but a less sound laboratory foreground. In such contexts the term *in theory, yes* is usually understood to be followed by the phrase *in practice, no.* Contrariwise, the experienced kineticist (even more so, the fresh graduate) may believe that the bench worker should use the term *conjecture* or the expression *set of vague and prejudiced hunches* rather than the fine word *theory* to describe the set of beliefs under which the latter is laboring.

The effects of *some* factors on *one* property of an industrial product may well be broadly guessed or even precisely predicted from available theory. But no industrial product has only one property of interest. It must be stable and inexpensive and small and inodorous and easy to use, and so on, through a list of perhaps 20 attributes. For many of these, little or no theory will be available. Even when theoretical methods might yield correct answers, it may be that no one is available who can use these methods expeditiously. Time will often be saved by simply "getting the data."

Most of my own experience with industrial experimentation has been near the empirical end of the spectrum just indicated, and this bias will show repeatedly in later chapters. The two-level multifactor fractional replicates—and other incomplete two-level factorials—which are one of the principal subjects of this work are quite surely of wide application when a broad range of experience must be accumulated economically in the absence of easily applied theory. Little, but still something, will be said about the prospects for other, more theoretically developed branches of industrial research.

Real differences of opinion on how best to proceed may become very important. Theoreticians may judge that a problem should first be studied "on paper"; laboratory workers may feel certain that the primary need is for more data. Compromises should be considered. Perhaps both views can be implemented at the same time. If the theoreticians can tell the laboratory workers what data they would most like to have, the information may be produced more quickly than either group thought possible. This is so because more can be found out *per run made* or per compound synthesized or per product modification carried out than most experimenters realize.

1.2. SCIENTIFIC METHODS

The research worker is often able to see the results of one run or trial before making another. He may guess that he can improve his yield, say, by a slight increase in temperature, by a considerable increase in pressure, by using a little more emulsifier, *and* adding a little more catalyst. He will act on all four guesses at once in his next run. And yet, in conversation, especially in general or philosophical conversation, he may state his belief in the value of varying *one factor at a time*. Indeed many experimenters identify the one-factor-at-a-time approach as "the" scientific method of experimentation.

Two different phases of research are being confused here. In the very early stages of any problem, *operability* experiments must be done, to see whether *any* yield or other desired property is attainable. After some set of operable or promising conditions has been established, the experimenter is very likely to continue trying simultaneous variation of all factors he thinks may help.

When this no longer works, he may well decide that he must settle down and vary the experimental conditions one at a time, in sequences that are natural for the system in question. This will often involve series of small increments for each of the continuously variable factors. Confusion appears when methods that seem appropriate for the later stage are claimed as valid for the earlier one.

As the process or product gets closer to the market, more and more *conditions and tolerances* turn up as requirements. Toxicity, inflammability, shelf life, and compatibility with dozens of other materials may have to be studied. The tolerance of the product to a wide variety of circumstance of use begins to assume major importance. The research or development technician must now investigate a whole set of new conditions. He must be able to assure the producing and marketing divisions of his company that the product can be guaranteed safe, efficient, and operable under a range of conditions not studied when it was first being considered and developed.

Because of the shortage of available technicians, because of the entire lack of any theory for some properties, because of the multiplicity of factors that *may* influence a product, and because of the other multiplicity of factors to which it must be insensitive, industrial research often differs widely from pure or basic research. In particular, more factors must be studied, and so it is often said, and rightly, that more data must be taken in industrial research problems than in pure research ones.

1.3. MAKING EACH PIECE OF DATA WORK TWICE*

It does not follow that the enormous amounts of data often accumulated in industrial research laboratories are entirely justified. Most experimenters, and most research directors too, I believe, have assumed that each piece of data can be expected to give information on the effect of one factor *at most*. This entirely erroneous notion is so widespread and so little questioned that its correction should start right here with the simplest possible example to the contrary.

A chemist has two small objects to weigh. He has a double-pan scale of fair precision and of negligible bias and a set of weights with excellent calibration. He would like to know the weight of each object with the best precision possible. He is to make *two* weighings only. His experience, habits, and common sense conspire to tell him to weigh one object (call it P) and then to weigh the other, Q—carefully of course. For each object there will be one weighing, one piece of data, one weight.

There is, however, a way to find the weight of each object as precisely as if *it* had been weighed twice and the two weighings averaged. To do this each

* This expression is due to W. J. Youden.

object must indeed be weighed twice. But we are allowed only two weighings in all. Hence each object must be on the scale pans twice. If the two objects are put in one pan and weighed together, we get an estimate of the sum of the two weights. To separate the components we must either weigh just one, or else find their difference. By placing one object in one pan and one in the other, we can, by balancing with the calibrated weights, find the difference. Calling the sum of the weights $S = P + Q$, and the difference $D = P - Q$, we see that the average of S and D measures the weight of P only, since Q is exactly balanced out. Similarly, the average of S and $-D$ measures the weight of Q with P exactly balanced out. We have then weighed each object twice, in two weighings, each with the precision of two averaged weighings.

The disadvantage of this "weighing design" is that no information is available until all the data are in. The reward for the delay is, in this case, the double precision. The moral, to be given extended emphasis and development later, is that each observation can be made to yield information on two (or more) parameters. Indeed the number of times that each observation can be used increases steadily with the number of observations in each balanced set. What is required is *planning*. In most cases, little or no information is extractable along the way. Finally a computation, usually quite simple, must be made to extract all the information at once.

The pronounced improvement of the (S, D) pair of weighings over the (P, Q) set becomes a minor matter when compared with the gains that are attainable when larger sets of weights or any other measurements are to be estimated. The simplest case was used here as an example that does not appear to have been mentioned since it was first pointed out by Hotelling [1942].

1.4. FIRST STAGES IN PLANNING INDUSTRIAL EXPERIMENTS

The stated aims of an industrial experiment are not the same at all of its stages, but the same broad desiderata seem to emerge repeatedly. We always want to know whether an effect holds fairly generally, and whether an apparent lack of effect of some factor is a general lack. Fisher's determined emphasis on the importance of a broad base for scientific inferences can never be forgotten. It is not a counsel of perfection but rather a sine qua non for good industrial research.

Some experimenters believe that they must be able to judge *early* which factors are going to be influential. They foresee, or think they do, that the experimental program will become unmanageably large if all factors are admitted for detailed study. But if factors are dropped from the active list too early, on the basis of relatively small numbers of data, it may take the research worker a long time to get back on the right track.

It is better to write down quite early a full list of all the factors that might influence the desired properties of the product under development. A valuable exercise in planning an attack on a new problem is to prepare a cross tabulation of all potential factors by all interesting properties of the product or process. This "influence matrix" should be a dated record of the experimenter's opinions about the effect(s) of each independently controllable variable on each property. Its use is discussed in Chapter 9.

Within the limits of practicability it is desirable to look at each factor's effects under a wide range of conditions or levels of the other factors. A stable effect, even at zero, over a wide range of settings of the other factors is reassuring because broadly based. On the other hand, if the effect of some factor varies, perhaps even changes sign depending on the settings of the others, this information is important and should be known early. Balanced or nearly balanced sets of runs provide the easiest way to learn about these situations.

Perhaps the major departure of this work from others with similar subject is its attitude toward the assumptions that are usually made before experimentation is started. The standard assumptions of most statistical treatments are as follows:

1. The observations must be a fair (representative, random) sample of the population about which inferences are desired.
2. The observations are of constant variance (or at least the variance must be a known function of the independent variables), are statistically independent, and are normally distributed.
3. Few or no bad values will be produced, and few missing values.

Assumption 1 is for the experimenters to guarantee. The three parts of assumption 2 can often be verified, or at least refuted, by the data themselves. Responding to the myriad ways in which data fail to meet these requirements will be a major part of the effort. Assumption 3 is violated in a large number, perhaps 30%, of all industrial experiments. Methods are given for spotting bad values, and for drawing valid conclusions, though often with reduced precision, in spite of these defects.

1.5. STATISTICAL BACKGROUND REQUIRED

I assume that the research worker reading this book knows a few of the fundamentals of applied statistics. Foremost among these are the following:

1. The prime requirement for drawing any valid inference from experimental data is that the inferrer know something about the way in which

the data (a sample) represent nature (the population). The prime requirement for the validity of any conclusions drawn from the study of experimental data is that the data be a real sample of the situations to which the conclusions are to apply.

2. The basic terms—*statistic, parameter, sample mean, sample standard deviation, standard error of a mean, regression coefficient, least-squares estimate*—should all be familiar. They will all be defined and described, but if the reader is encountering many of them for the first time, he will not find these pages easy reading.
3. The most pervasive generalization in the whole of statistics is the Central Limit Theorem. Its effect is to make averages of independent observations more nearly Gaussian in their distribution than the error distributions of the single observations. Since a large proportion of the parameter estimates we make are averages, the central limit theorem must be working for us a large part of the time. This comforting circumstance cannot account for the apparent "normality" we will repeatedly find in residuals, however, since they are heavily dependent on the single observations themselves. For these we must believe that a considerable number of small additive, nearly independent factors are responsible. No quantitative knowledge or application of the theorem is ever necessary. It simply operates, like a law of nature, but, unlike other laws, generally in our favor. The reader is referred to Cramér [1946] for an illuminating discussion of the central limit theorem and of its antecedents.

1.6. DOING THE ARITHMETIC

Many research engineers and industrial scientists are repelled by the monotonous and extensive arithmetic that statistical texts and handbooks seem to demand. My sympathies are with them; much of this drudgery is unnecessary. Nearly all the arithmetic in this book has been done by hand, perhaps on a desk calculator. Intelligent coding and rounding are of the essence and frequently result in reducing time, as well as errors, to a small fraction of their former magnitudes.

When 10 or more experiments (or responses in a single experiment of size 16 or larger) must be analyzed, time will be saved if the standard algorithms (for the analysis of variance, for Yates's method in 2^n plans, for partially balanced incomplete blocks) are available on a computer. *Do not consider any program that does not compute and print residuals automatically, preferably rounded to two digits.*

The plotting of cumulative empirical distributions of residuals on a "normal" grid is again a tedious job when done as proposed in the few textbooks that mention it. But when the number of points is large, the job can be greatly shortened. Standard grids for $N = 16, 32$ are given that require no calculation of probability points. All large computers have been programmed to a fare-thee-well to make approximate plots without special peripheral equipment, and only approximate plots are needed. I have found too that, when the number of points exceeds 100, it is usually necessary to plot only the largest 25 or so (including both ends). As soon as the plotted set "point" straight through the 50% point, there is no need to continue plotting.

1.7. SEQUENCES OF EXPERIMENTS

The *analysis* of sequences of agricultural experiments has been studied extensively by Yates and Cochran [1957, pages 565 ff.], and much can be learned from this work. The *design* of sequences of industrial experiments is much less fully developed, although economical augmentation of early experiments seems to be crucial in industrial research. The earliest work in this area was by Davies and Hay [1950]. Less clear, but more economical, augmentations were published in 1962 [Daniel]. [Although trend-robust plans (Chapter 15) are carried out in sequence, they are not really adaptive designs but have to be carried all the way before effects can be estimated.]

1.8. THE FUTURE OF THE DESIGN OF INDUSTRIAL EXPERIMENTS

Major new developments in the design of industrial experiments seem to me to await the appearance of well-educated statisticians who want to work in close touch with industrial scientists. Many mathematical statisticians are under the illusion that they and their graduate students are writing for a future which they forsee without benefit of detailed knowledge of the present. A tiny proportion of their work may be remembered 20 years from now.

As in the past, many developments will come from scientists and engineers with extensive experience in industrial research. But we need in addition a cohort of modest graduate statisticians who recognize the productiveness of going directly to industrial scientists to find out just how they do their research. Far too many graduates, and even some senior statisticians, are willing if not anxious to tell scientists how to plan their experiments, in advance of knowing just how such work is now done. There are, fortunately, a few outstanding exceptions. I think especially of the work of Box, Lucas, Behnken, W. G. Hunter, N. R. Draper, and their associates on "nonlinear" design. A shortcoming of this book is its lack of any treatment of these plans—an omission due to my own lack of experience with them.

This Page Intentionally Left Blank

CHAPTER 2

Simple Comparison Experiments

2.1. AN EXAMPLE

The example is taken from Davies [1956, Ed. 2, printing 1971, pages 12–18]. Only one criticism is to be made, and that with some hesitation, since this is the fundamental work on industrial experimentation (from which I for one have learned more than from any other book).

We quote from Davies, Section 2.21:

> The experiment was required to test whether or not treatment with a certain chlorinating agent increased the abrasion resistance of a particular type of rubber. The experimenter took ten test-pieces of the material and divided each piece into two. One half was treated and the other half was left untreated, the choice of which half of the specimen should receive the treatment being made by tossing a coin. The abrasion resistances of the ten pairs of specimens were then tested by a machine, the specimens being taken in random order.

Perhaps most experimenters would prefer to call such a collection of data a *test*, so as not to invoke the grander connotations of the term *scientific experiment*. It is not clear from the description or from later discussion (page 43, Figure 2.5) whether all 10 specimens were taken from one sheet of rubber. Since we need a straw man for this discussion, let us assume that the 10 were indeed a random sample from a single sheet. Randomization of the choice of half piece for chlorination plus random allocation of sample points in the sheet of rubber have guaranteed that any difference found and judged to be real has a good chance of being confirmed if measured over the whole sheet.

But the data come from one sheet of rubber. The pains taken to obtain precise and "unbiased" data have resulted in our getting into our sample