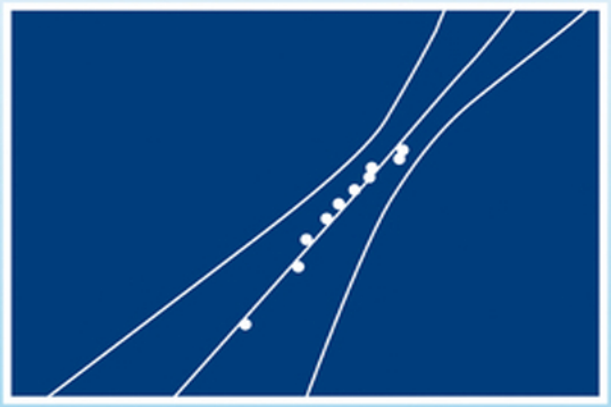
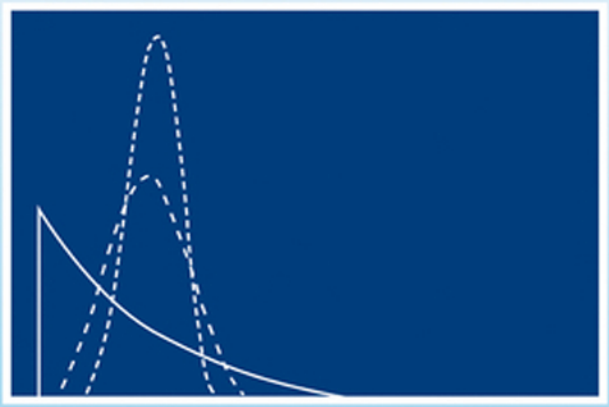
Statistics and Probability with Applications for Engineers and Scientists





Bhisham C. Gupta • Irwin Guttman

WWW. LINK AVAILABLE

WILEY

STATISTICS AND PROBABILITY WITH APPLICATIONS FOR ENGINEERS AND SCIENTISTS

STATISTICS AND PROBABILITY WITH APPLICATIONS FOR ENGINEERS AND SCIENTISTS

Bhisham C. Gupta

Professor of Statistics
University of Southern Maine
Portland, ME

Irwin Guttman

Professor Emeritus of Statistics
SUNY at Buffalo and
University of Toronto, Canada

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Gupta, Bhisham C., 1942-
Statistics and probability with applications for engineers and scientists / Bhisham C. Gupta, Department of Mathematics and Statistics, University of Southern Maine, Portland, ME, Irwin Guttman, Amherst, NY.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-46404-5 (hardback)
1. Probabilities. 2. Mathematical statistics. I. Guttman, Irwin. II. Title.
QA273.G85 2013
519.5–dc23

2012032883

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

In the loving memory of my parents, Roshan Lal and Sodhan Devi
–Bhisham

In the loving memory of my parents, Anna and Samuel Guttman
–Irwin

Modern statisticians are familiar with the notion that any finite body of data contains only a limited amount of information on any point under examination; that this limit is set by the nature of the data themselves, and cannot be increased by any amount of ingenuity expended in their statistical examination: that the statistician's task, in fact, is limited to the extraction of the whole of the available information on any particular issue.

R. A. Fisher

Table of Contents

Preface

AUDIENCE

This is an introductory textbook in applied statistics and probability for undergraduate students in engineering and the natural sciences. It begins at a level suitable for those with no previous exposure to probability and statistics and carries the reader through to a level of proficiency in various techniques of statistics. This text is divided into two parts: Part I discusses descriptive statistics, concepts of probability, probability distributions, sampling distributions, estimation, and testing of hypotheses, and Part II discusses various topics of applied statistics, including some reliability theory, statistical quality control charts of phase I and phase II, some nonparametric techniques, categorical data analysis, simple and multiple linear regression analysis, design and analysis of variance with emphasis on 2^k factorial designs, and response surface methodology.

This text is suitable for a one- or two-semester undergraduate course sequence. The presentation of material gives instructors a lot of flexibility to pick and choose topics they feel should make up the coverage of material for their courses. However, we feel that in the first course for engineers and science majors, one may cover Chapters 1 and 2, a brief discussion of probability in Chapter 3, selected discrete and continuous distributions from Chapters 4 and 5 with more emphasis on normal distribution, Chapters 7 through 9, and couple of topics from Part II that meet the needs and interests of the particular group of students. For example, some discussion of the material on regression analysis and design of experiments in Chapters 15 and 17 may serve well. A two-semester course may cover the entire book. The only prerequisite is a first course in calculus, which all engineering and science students are required to take.

Because of space considerations, some proofs and derivations, certain advance level topics of interest, including Chapter 19 on response surfaces, are not included in the text, but are available for download via the book's website: www.wiley.com/go/statsforengineers.

MOTIVATION

Students encounter data-analysis problems in many areas of engineering or natural science curricula. Engineers and scientists in their professional lives often encounter situations requiring analysis of data arising from their areas of practice. Very often they have to plan the investigation that generates data (an activity euphemistically called the design of experiments), analyze the data obtained, and interpret the results. Other problems and investigations may pertain to the maintenance of quality of existing products or the development of new products, or to a desired outcome in an investigation of the underlying mechanisms governing a certain process.

Knowing how to "design" a particular investigation to obtain reliable data must be coupled with knowledge of descriptive and inferential statistical tools to analyze properly and interpret such data. The intent of this textbook is to expose the uninitiated to statistical methods that deal with the generation of data for different (but frequently met) types of investigations and to discuss how to analyze and interpret the generated data.

HISTORY

This text has its roots in the three editions of *Introductory Engineering Statistics*, first co-authored by Irwin Guttman and the late, great Samuel Wilks. Professor J. Stuart Hunter (Princeton University), one of the finest expositors in the statistics profession, a noted researcher, and a colleague of Professor Wilks, joined Professor Guttman to produce editions two and three. All editions were published by John Wiley & Sons, with the third edition appearing in 1982.

APPROACH

In this text we emphasize both descriptive and inferential statistics. We first give details of descriptive statistics and then continue with an elementary discussion of the fundamentals of probability theory underlying many of the statistical techniques discussed in this text. We next cover a wide range of statistical techniques such as statistical estimation, regression methods, statistical quality control (with emphasis on phase I and phase II control charts), and process capability indices, nonparametric methods, elements of reliability theory, and the like. A feature of these discussions is that all statistical concepts are supported by a large number of examples using data encountered in real-life situations. We also illustrate how the statistical packages MINITAB® Version 16, Microsoft Excel® Version windows 2007, and JMP® Version 9, may be used to aid in the analysis of various data sets.

Another feature of this text is the coverage at an adequate and understandable level of the design of experiments. This includes a discussion of randomized block designs, one- and two-way designs, Latin square designs, 2^k factorial designs, response surface designs, among others. As indicated above, all this is illustrated with real-life situations and accompanying data sets, supported by MINITAB, Microsoft Excel, and JMP. We know of no other book in the market that covers all these software packages.

HALLMARK FEATURES

Software Integration

As indicated above, we incorporate MINITAB and Microsoft Excel throughout the text and present JMP at the end of each chapter. Our step-by-step approach to the use of the software packages means no prior knowledge of their use is required. After completing a course that uses this text, students will be able to use these software packages to analyze statistical data in their fields of interest.

Breadth of Coverage

Besides the coverage of many popular statistical techniques, we include discussion of certain aspects of sampling distributions, nonparametric tests, phase II control charts, reliability theory, design of experiments, and response surface methodology. Phase II control charts are discussed in a separate chapter that includes the use of the statistical packages to implement these charts.

Design of experiments and response surface methodology are treated in sufficient breadth and depth to be appropriate for a two-course sequence in engineering statistics that includes probability and the design of experiments.

Real data in examples and homework problems illustrate the importance of statistics and probability as a tool for engineers and scientists in their professional lives. All the data sets with 20 or more data points are available on the website in three formats: MINITAB, Microsoft Excel, and JMP.

Case studies in each chapter further illustrate the importance of statistical techniques in professional practice.

STUDENT RESOURCES

Data sets for all examples and homework exercises from the text are available to students on the website in Minitab, Microsoft Excel, and JMP format. The sample data sets were generated using well-known statistical sampling procedures, ensuring that we are dealing with *random samples*. An inkling of what this may entail is given throughout the text. (See, for example, Section 7.1.2.) The field of sampling is an active topic among research statisticians and practitioners, and references to sampling techniques are widely available in books and journal articles. Some of these references are included in the bibliography section.

Other resources on the book website www.wiley.com/go/statsforengineers available for download include:

Solutions Manual to all odd numbered homework exercises in the text.

Excel worksheets and macros to supplement features that are not available in Excel.

INSTRUCTOR RESOURCES

The following resources are available to adopting instructors on the textbook website www.wiley.com/go/statsforengineers:

Solutions Manual to all homework exercises in the text.

Lecture slides to aid instructors preparing for lectures.

Data sets for all examples and homework exercises from the book, in three formats: Minitab, Microsoft Excel, and JMP.

Excel worksheets and macros to supplement features that are not available in Excel.

Errata

We have thoroughly reviewed the text to make sure it is as error-free as possible. However, any errors discovered will be listed on the textbook website.

If you encounter any errors as you are using the book, please send them directly to the authors (bcgupta@usm.maine.edu), so that the errors can be corrected in a timely manner on the website, and for future editions. We also welcome any suggestions for improvement you may have, and thank you in advance for helping us improve the book for future readers.

ACKNOWLEDGMENTS

We are grateful to the following reviewers and colleagues whose comments and suggestions were invaluable in improving the text:

Zaid Abdo, University of Idaho
Erin Baker, University of Massachusetts
Bob Barnet, University of Wisconsin-Platteville
Mark Gebert, University of Kentucky
Ramesh Gupta, University of Maine
Rameshwar Gupta, University of New Brunswick, Canada
Xiaochun Jiang, North Carolina Agricultural and Technical State University
Dennis Johnston, Baylor University
Gerald Keller, Joseph L. Rotman School of Management, University of Toronto
Kyungduk Ko, Boise State University

Paul Kvam, Georgia Institute of Technology
Thunshun Liao, Louisiana State University
Jye-Chyi Lu, Georgia Institute of Technology
Sumona Mondal, Clarkson University
Janbiao Pan, California Poly State University
Anastassios Perakis, University of Michigan
David Powers, Clarkson University
Ali Touran, Northeastern University
Leigh Williams, Virginia Polytechnic and State University
Tian Zheng, Columbia University
Jingyi Zhu, University of Utah

We thank William Belcher, Darwin Davis, Julie Ellis, Pushpa Gupta, Mohamad Ibourk, James Lucas, Mary McShane-Vaughn, Louis Neveux, and Phil Ramsey who helped find suitable data sets for the case studies. We also thank Laurie McDermott for her help in typing some parts of this manuscript. Special thanks are due to Eric Laflamme for helping write JMP/Excel procedures and creating PowerPoint® presentations, George Bernier for helping write Excel work books and macros, and Patricia Miller and Brenda Townsend for editing Power Point Slides and some parts of the manuscript.

We acknowledge Minitab Inc., SAS Institute Inc., and Microsoft for permitting us to print MINITAB, JMP, and Microsoft Excel screen shots in this book.

Portions of the text are reproduced by permission of the American Society for Quality (ASQ), *Applied Statistics for the Six Sigma Green Belt* and *Statistical Quality Control for the Six Sigma Green Belt* by Bhisham C. Gupta and H. Fred Walker (Milwaukee: ASQ Quality Press, 2005, 2007). To order these books, call ASQ at 800-248-1946 or 414-272-8575, or visit http://www.asq.org/quality-press.

We would also like to express our thanks and appreciation to the individuals at John Wiley, for their support, confidence, and guidance as we have worked together to develop this project.

The authors would like to gratefully thank their families. Bhisham acknowledges the patience and support of his wife, Swarn; daughters, Anita and Anjali; son, Shiva; sons-in-laws, Prajay and Mark; daughter-in-law, Aditi; and wonderful grandchildren, Priya, Kaviya, and Ayush and Amari. For their patience and support, Irwin is grateful to his wife, Mary; son, Daniel; daughters, Karen and Shaun; wonderful grandchildren, Liam, Teia, and Sebastian; brothers and their better halves, Alvin and Rita, and Stanley and Gloria.

BHISHAM GUPTA
IRWIN GUTTMAN

1 Introduction

Statistics, the discipline, is the study of the scientific method. In pursuing this discipline, Statisticians have developed a set of techniques that are extensively used to solve problems in any field of scientific endeavor, such as in the engineering sciences, biological sciences, and the chemical, pharmaceutical, and social sciences.

This book is concerned with discussing these techniques and their applications for certain experimental situations. It begins at a level suitable for those with no previous exposure to probability and statistics, and carries the reader through to a level of proficiency in various techniques of statistics.

In all scientific areas, whether engineering, biological sciences, medicine, chemical, pharmaceutical, or social sciences, scientists are inevitably confronted with problems that need to be investigated. Consider some examples:

- An engineer wants to determine the role of an electronic component needed to detect the malfunction of the engine of a plane.
- A biologist wants to study various aspects of wildlife, the origin of a disease, or the genetic aspects of a wild animal.
- A medical researcher is interested in determining the cause of a certain type of cancer.
- A manufacturer of lenses wants to study the quality of the finishing on intraocular lenses.
- A chemist is interested in determining the effect of a catalyst in the production of low-density polyethylene.
- A pharmaceutical company is interested in developing a vaccination for swine flu.
- A social scientist is interested in exploring a particular aspect of human society.

In all of the examples above, the first and foremost work is to define clearly the objective of the study and precisely formulate the problem. The next important step is to gather information to help determine what key factors are affecting the problem. Remember that to determine these factors successfully, you should understand not merely statistical methodology but relevant nonstatistical knowledge as well. Once the problem is formulated and the key factors of the problem are identified, the next step is to

Statistics and Probability with Applications for Engineers and Scientists, Bhisham C. Gupta and Irwin Guttman.

collect the data. There are various methods of data collecting. Four basic methods of statistical data collecting are:

- A designed experiment
- A survey
- An observational study
- A set of historical data, that is, data collected by an organization or an individual in an earlier study

1.1 DESIGNED EXPERIMENT

We discuss the concept of a designed experiment with an example, "Development of Screening Facility for Storm Water Overflows" (taken from Box, Hunter, and Hunter, 1978, and used with permission). The example illustrates how a sequence of experiments can enable scientists to gain knowledge of the various *important factors* affecting the problem, and give insight into the objectives of the investigation. It also indicates how unexpected features of the problem can become dominant, and how experimental difficulties can occur so that certain planned experiments cannot be run at all. Most of all, this example shows the importance of common sense in the conduct of any experimental investigation. The reader may rightly conclude from this example that the course of a real investigation, like that of true love, seldom runs smoothly, although the eventual outcome may be satisfactory.

1.1.1 Motivation for the Study

During heavy rainstorms, the total flow coming to a sewage treatment plant may exceed its capacity, making it necessary to bypass the excess flow around the treatment plant, as shown in Figure 1.1.1a. Unfortunately, the storm overflow of untreated sewage causes pollution of the receiving body of water. A possible alternative, sketched in Figure 1.1.1b, is to screen most of the solids out of the overflow in some way and return them to the plant for treatment. Only the less objectionable screened overflow is discharged directly to the river.

To determine whether it was economical to construct and operate such a screening facility, the Federal Water Pollution Control Administration of the Department of the Interior sponsored a research project at the Sullivan Gulch pump station in Portland, Oregon. Usually the flow to the pump station was 20 million gallons per day (mgd), but during a storm, the flow could exceed 50 mgd.

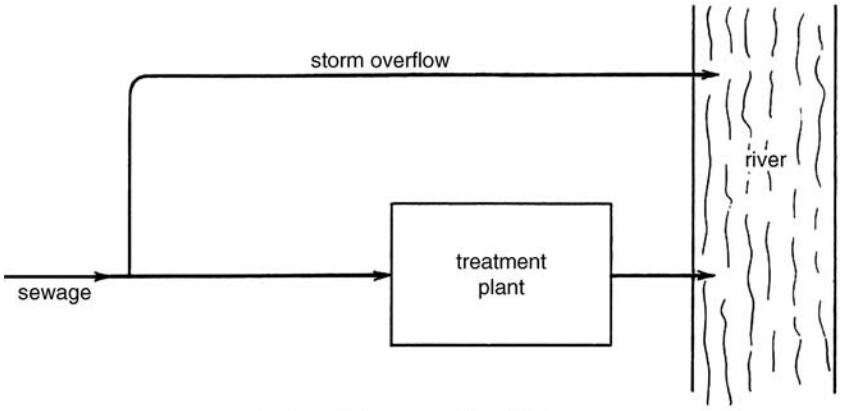
Figure 1.1.2a shows the original version of the experimental screening unit, which could handle approximately 1000 gallons per minute (gpm). Figure 1.1.2a is a perspective view and Figure 1.1.2b is a simplified schematic diagram. A single unit was about seven feet high and seven feet in diameter. The flow of raw sewage struck a rotating collar screen at a velocity of five to 15 feet per second. This speed was a function of the flow rate into the unit and hence a function of the diameter of the influent pipe. Depending on the speed of the rotation of this screen and its fineness, up to 90% of the feed penetrated the collar screen. The rest of the feed dropped to the horizontal screen, which vibrated to remove excess water. The solids concentrate, which passed through neither screen, was sent to the sewage treatment plant. Unfortunately, during operation the screens became clogged with solid matter, not only sewage but also oil, paint, and fish-packing wastes. Backwash sprays were therefore installed for both screens to permit cleaning during operation.

1.1.2 Investigation

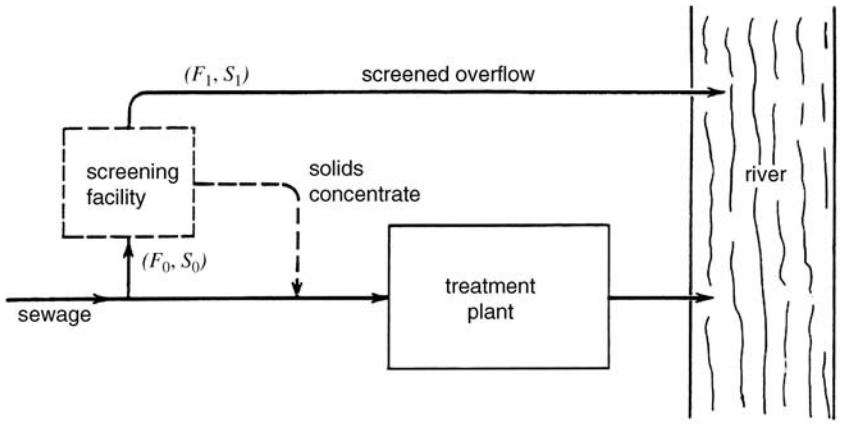
The objective of the investigation was to determine good operating conditions.

1.1.3 Changing Criteria

What are good operating conditions? Initially it was believed they were those resulting in the highest possible removal of solids. In Figure 1.1.1b, settleable solids in the influent are denoted by S_0 and



(*a*) Standard mode of operation.



(*b*) Modified mode of operation, with screening facility. F = flow, S = settleable solids.

FIGURE 1.1.1 Operation of the sewage treatment plant.

the settleable solids in the effluent by S_1. The *percent solids removed* by the screen is therefore $y = 100(S_0 - S_1)/S_0$. Thus, initially, it was believed that good operation meant achieving a high value for y. However, it became evident after the first set of experimental made that the *percentage of the flow retreated* (flow returned to treatment plant), which we denote by z, also had to be taken into account. In Figure 1.1.1b, influent flow to the screens is denoted by F_0 and effluent flow from the screens to the river by F_1. Thus $z = 100(F_0 - F_1)/F_0$.

1.1.4 A Summary of the Various Phases of the Investigation

Phase a

In this initial phase an experiment was run in which the roles of three variables were studied: collar screen mesh size (fine, coarse), horizontal screen mesh size (fine, coarse), and flow rate (gallons per minute). At this stage:

1. The experimenters were encouraged by the generally high values achieved for y.
2. Highest values for y were apparently achieved by using a horizontal screen with a coarse mesh and a collar screen with fine mesh.

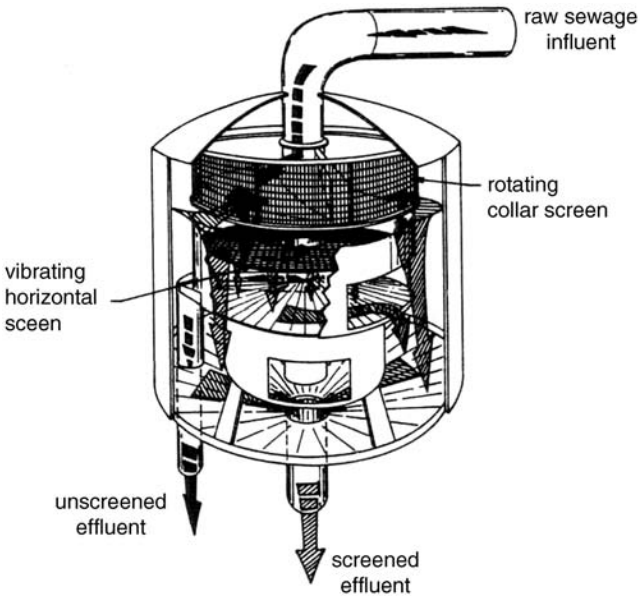


FIGURE 1.1.2a Original version of the screening unit (detailed diagram).

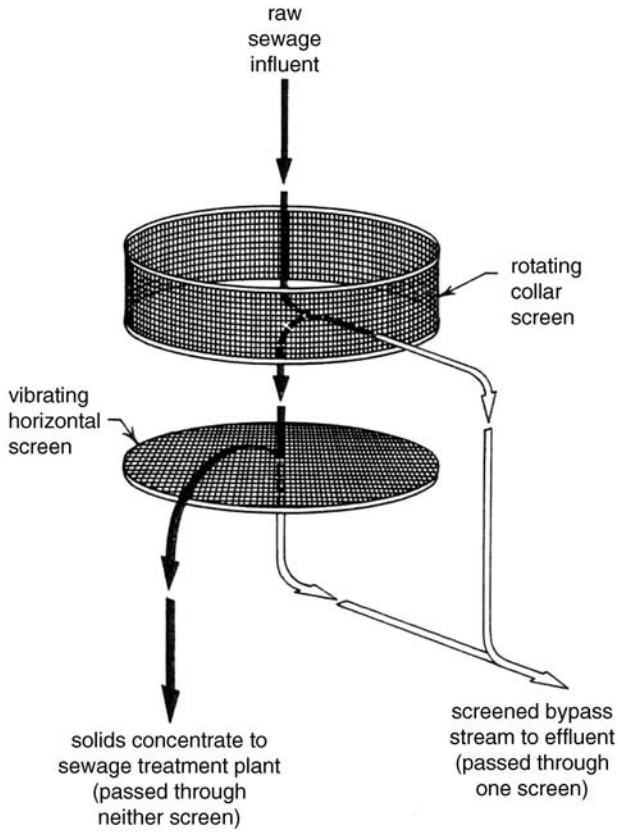


FIGURE 1.1.2b Original version of the screening unit (simplified diagram).

3. Contrary to expectation, flow rate did not show up as an important variable affecting y.
4. Most important, the experiment was unexpectedly dominated by the z values, which measure the flow retreated. These were uniformly very low, with about 0.01% of the flow being returned to the treatment plant and 99.9% leaving the screen for discharge into the river. Although it was desirable that the retreated flow be small, the z values were embarrassingly low. As the experimenters remarked, "[T]he horizontal screen produced a solid concentrate . . . dry enough to shovel . . . . This represented a waste of effort of concentrating because the concentrated solids were intended to *flow* from the units."

Phase b

It was now clear (1) that z as well as y was important and (2) that z was too low. It was conjectured that the matters might be improved by removing the horizontal screen altogether. Another experiment was therefore performed with no horizontal screen. The speed of rotation of the collar screen was introduced as a new variable.

Unfortunately, after only two runs of this experiment this particular phase had to be terminated because of the excessive tearing of the cloth screens. From the scanty results obtained it appeared, however, that with no horizontal screen high solids removal could be achieved with a higher portion of the flow retreated. It was therefore decided to repeat these runs with screens made of stainless steel instead of cloth.

Phase c

A third experiment, using stainless steel collar screens of two mesh sizes, similar to that attempted in phase b, was performed with the same collar screen mesh size, collar screen speed (rpm), and flow rate (gpm) used before.

In this phase with a stainless steel collar screen, high removal rates y were possible for eight sets of conditions for the factors just mentioned. However, these high y values were obtained with retreated flow z at undesirably high values (before, they had been too low). The object was to get reasonably small values for z, but not so small as to make shoveling necessary; values between 5% and 20% were desirable. It was believed that by varying flow rate and speed of rotation of the collar screen, this objective could be achieved without sacrificing solids removal.

Phase d

Again, using a stainless steel collar screen, another experiment, with two factors—collar screen speed (rpm) and flow rate (gpm)—set at two levels each, was run. This time, high values of solids removal were maintained, but unfortunately flow retreated values were even higher than before.

Phase e

It was now conjectured that intermittent back washing could overcome the difficulties. This procedure was now introduced with influent flow rate and collar screen mesh varied.

The results of this experiment lead to a removal efficiency of 89% with a retreated flow of only 8%. This was regarded as a satisfactory and practical solution, and the investigation was terminated at that point.

For detailed analysis of this experiment, the reader should refer to Box, Hunter, and Hunter (1978), p. 354. Of course, these types of experiments and their analyses are discussed in this text (see Chapter 18).

1.2 A SURVEY

The purpose of a sample survey is to make inferences about certain characteristics of a population from which samples are drawn. The inferences to be made for a population usually entails the estimation of population parameters, such as the population total, the mean, or the population proportion of a certain

characteristic of interest. In any sample survey a clear statement of its objective is very important. Without a clear statement about the objectives, it is very easy to miss pertinent information while planning the survey that can cause difficulties at the end of the study.

In any sample survey only relevant information should be collected. Sometimes trying to collect too much information may become very confusing and consequently hinder the determination of the final goal. Moreover, collecting information in sample surveys costs money, so that the interested party must determine which and how much information should be obtained. For example, it is important to describe how much precision in the final results is desired. Too little information may prevent obtaining good estimates with desired precision, while too much information may not be needed and may unnecessarily cost too much money. One way to avoid such problems is to select an appropriate method of sampling the population. In other words, the sample survey needs to be appropriately designed. A brief discussion of such designs is given in Chapter 2. For more details on these designs, the reader may refer to Cochran (1977), Sukhatme et al. (1970), or Schaeffer et al. (2006).

1.3 AN OBSERVATIONAL STUDY

An observational study is one that does not involve any experimental studies. Consequently observational studies do not control any variables. For example, a realtor wishes to appraise a house value. All the data used for this purpose is observational data. Many psychiatric studies involve observational data.

Frequently, in fitting a regression model (see Chapters 15 and 16), we use observational data. Similarly, in quality control (see Chapters 11 and 12), most of the data used in studying control charts for attributes is observational data. Note that control charts for attributes usually do not provide any cause-and-effect relationships. This is because observational data give us very limited information about cause-and-effect relationships.

As another example, many psychiatric studies involve observational data, and such data do not provide the cause of patient's psychiatric problems. An advantage of observational studies is that they are usually more cost effective than experimental studies. The disadvantage of observational studies is that the data may not be as informative as experimental data.

1.4 A SET OF HISTORICAL DATA

Historical data are not collected by the experimenter. The data are made available to him/her.

Many fields of study such as the many branches of business studies, use historical data. A financial advisor for planning purposes uses sets of historical data. Many investment services provide financial data on a company-by-company basis.

1.5 A BRIEF DESCRIPTION OF WHAT IS COVERED IN THIS BOOK

Data collection is very important since it can greatly influence the final outcome of subsequent data analyses. After collection of the data, it is important to organize, summarize, present the preliminary outcomes, and interpret them. Various types of tables and graphs that summarize the data are presented in Chapter 2. Also in that chapter we give some methods used to determine certain quantities, called *statistics*, which are used to summarize some of the key properties of the data.

The basic principles of probability are necessary to study various probability distributions. We present the basic principles of elementary probability theory in Chapter 3. Probability distributions are fundamental in the development of the various techniques of statistical inference. The concept of random variables is also discussed in Chapter 3.

Chapters 4 and 5 are devoted to some of the important discrete distributions, continuous distributions, and their moment-generating functions. In addition we study in Chapter 5 some special distributions that are used in reliability theory.

In Chapter 6 we study joint distributions of two or more discrete and continuous random variables and their moment-generating functions. Included in Chapter 6 is the study of the bivariate normal distribution.

Chapter 7 is devoted to the probability distributions of some sample statistics, such as the sample mean, sample proportions, and sample variance. In this chapter we also study a fundamental result of probability theory, known as the Central Limit Theorem. This theorem can be used to approximate the probability distribution of the sample mean when the sample size is large. In this chapter we also study some sampling distributions of some sample statistics for the special case in which the population distribution is the so-called normal distribution. In addition we present probability distributions of various "order statistics," such as the largest element in a sample, smallest element in a sample, and sample median.

Chapter 8 discusses the use of sample data for estimating the unknown population parameters of interest, such as the population mean, population variance, and population proportion. In Chapter 8 also discusses the methods of estimating the difference of two population means, the difference of two population proportions, and the ratio of two population variances and standard deviations. Two types of estimators are included, namely point estimators and interval estimators (confidence intervals).

Chapter 9 deals with the important statistical tests of hypotheses that are concerned with the population means, population variance, and population proportion for one and two populations. Methods of testing hypotheses using the confidence intervals studied in Chapter 8 are also presented.

Chapter 10 gives an introduction to the theory of reliability. Methods of estimation and hypothesis testing using the exponential and Weibull distributions are presented.

Chapters 11 and 12 are devoted to control charts for variables and attributes used in phase I and phase II of a process. "Phase I" refers to the initial stage of a new process, and "phase II" refers to a matured process. Control charts are used to determine whether a process involving manufacturing or service is "under statistical control" on the basis of information contained in a sequence of small samples of items of interest.

Chapter 13 is concerned with the chi-square goodness-of-fit test, which is used to test whether a set of sample data support the hypothesis that the sampled population follows some specified probability model. In addition we apply the chi-square goodness-of-fit test for testing hypotheses of independence and homogeneity. These tests involve methods of comparing observed frequencies with those that are expected if a certain hypothesis is true.

Chapter 14 gives a brief look at tests known as "nonparametric tests," which are used when the assumption about the underlying distribution having some specified parametric form cannot be made.

Chapter 15 introduces an important topic of applied statistics: simple linear regression analysis. Linear regression analysis is frequently used by engineers, social scientists, health researchers, and biological scientists. This statistical technique explores the relation between two variables so that one variable can be predicted from the other. In this chapter we discuss the least squares method for estimating the simple regression model, called the fitting of this regression model. Also we discuss how to perform a residual analysis, which is used to check the adequacy of the regression model, and study certain transformations that are used when the model is not adequate.

Chapter 16 extends the results of Chapter 15 to multiple linear regression. Like the simple linear regression model, multiple linear regression analysis is widely used. It provides statistical techniques that explore the relations among more than two variables, so that one variable can be predicted from the use of the other variables. In this chapter we give a discussion of multiple linear regression, including the matrix approach. Finally, a brief discussion of logistic regression is given.

In Chapter 17 we introduce the design and analysis of experiments using one, two, or more factors. Designs for eliminating the effects of one or two nuisance variables along with a method of estimating one or more missing observations are given. We include two nonparametric tests, the Kruskal–Wallis and the Friedman test, for analyzing one-way and randomized complete block designs. Finally, models with fixed effects, mixed effects, and random effects are also discussed.

Chapter 18 introduces a special class of designs, the so-called 2^k factorial designs. These designs are widely used in various industrial and scientific applications. An extensive discussion of unreplicated 2^k factorial designs, blocking of 2^k factorial designs, confounding in the 2^k factorial designs, and Yates's

algorithm for the 2^k factorial designs is also included. We also devote a section to fractional factorial designs, discussing one-half and one-quarter replications of 2^k factorial designs.

In Chapter 19 we introduce the topic of response surface methodology (RSM). First-order and second-order designs used in response surface methodology are discussed. Methods of determining optimum or near optimum points using the "method of steepest ascent" and the analysis of a fitted second-order response surface are also presented. Due to lack of space, this chapter is not included in the text but is available for download on the book website: www.wiley.com/go/statsforengineers.

All chapters are supported by three popular statistical software packages, MINITAB, Microsoft Excel, and JMP. The MINITAB and Microsoft Excel are fully integrated into the text of each chapter, whereas JMP is given in an independent section, which is not included in the text but is available for download on the book website: www.wiley.com/go/statsforengineers. Frequently we use the same examples for the discussion of JMP as are used in the discussion of MINITAB and Microsoft Excel. For the use of each of these software packages, no prior knowledge is assumed, since we give each step, from entering the data to the final analysis of such data under investigation. Finally, a section of case studies is included in almost all the chapters.