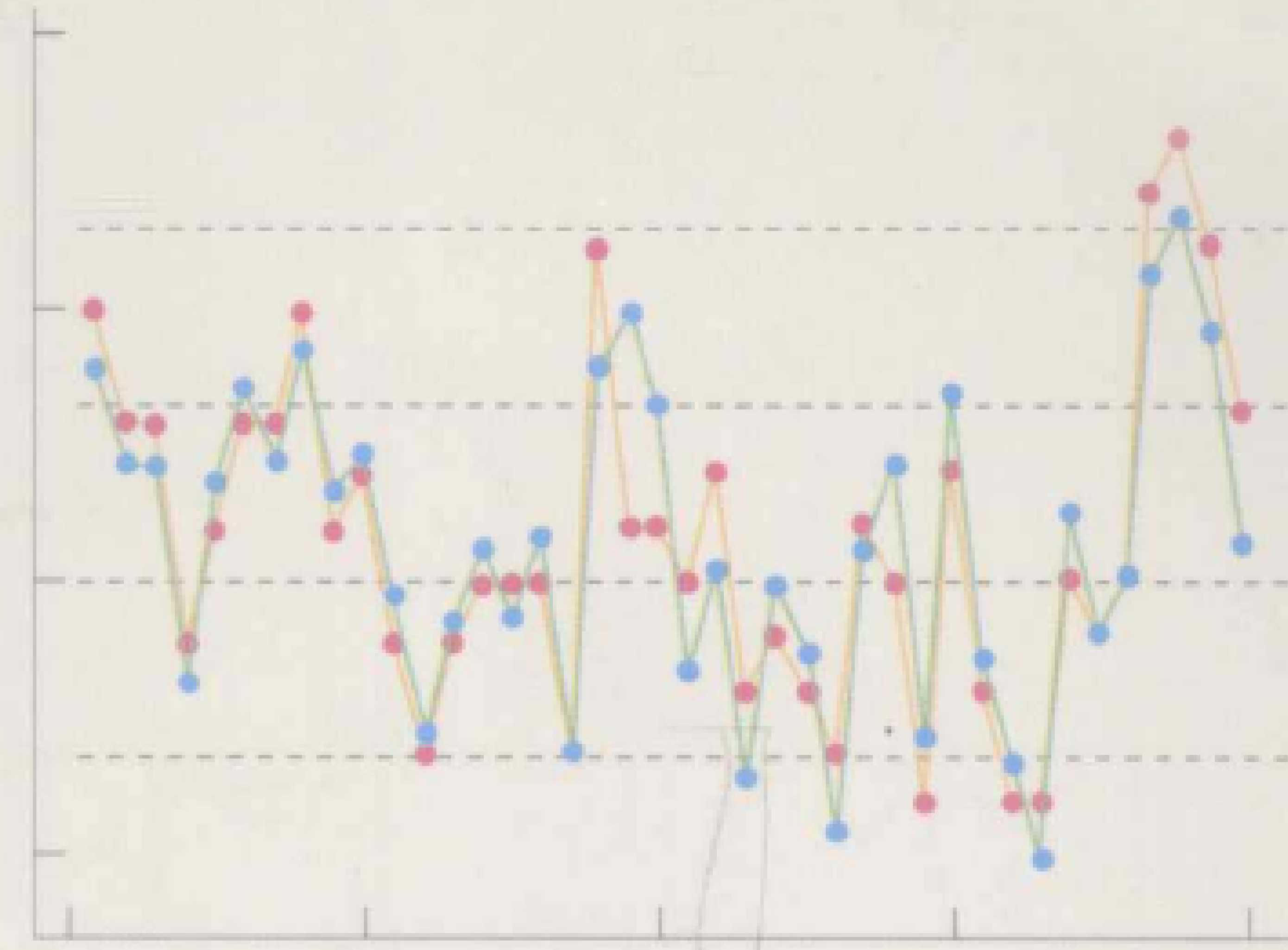
高等学校教材系列



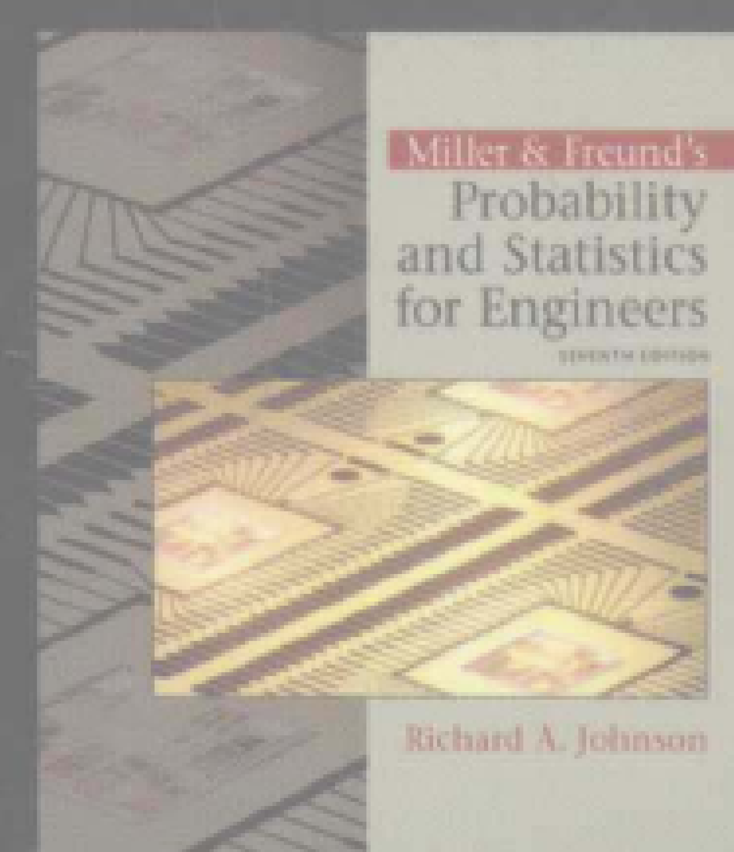
概率论与数理统计

（第七版）（英文版）

Miller & Freund's

Probability and Statistics for Engineers

Seventh Edition



[美] Richard A. Johnson 著
章栋恩 改编

電子工業出版社
Publishing House of Electronics Industry
http://www.phei.com.cn

高等学校教材系列

概率论与数理统计

（英文版）（第七版）

Miller & Freund's
Probability and Statistics for Engineers

Seventh Edition

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内容简介

这是一本为工程、物理等专业大学低年级学生准备的入门级教材。本书讲述了统计分布、概率密度、抽样分布、均值、方差、曲线拟合等概率论与数理统计中涉及到的基本概念。其特点是内容全面、编排合理，各章之间的衔接较为密切，附录中还提供了大量的常用统计表。全书用丰富的例子详细说明每一个概念，而省略了不必要的理论证明，例子中的数据大多来自作者的亲身经历及实际的案例。

本书适合作为大学工程、物理等专业一二年级学生的教材，也适合作为工程技术人员的参考材料。

版权贸易合同登记号　图字：01-2005-5659

图书在版编目（CIP）数据

概率论与数理统计 = Miller & Freund's Probability and Statistics for Engineers：第7版/（美）约翰逊（Johnson, R. A.）著；章栋恩改编. －北京：电子工业出版社，2005.12

（高等学校教材系列）

ISBN 7-121-01931-0

Ⅰ. 概… Ⅱ. ①约… ②章… Ⅲ. ①概率论－高等学校－教材－英文 ②数理统计－高等学校－教材－英文 Ⅳ. O21

中国版本图书馆 CIP 数据核字（2005）第 131048 号

责任编辑：许菊芳

印　　刷：北京市天竺颖华印刷厂

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱　邮编：100036

经　　销：各地新华书店

开　　本：787 × 980　1/16　　印张：36.25　　字数：812 千字

印　　次：2005 年 12 月第 1 次印刷

定　　价：58.00 元

凡购买电子工业出版社的图书，如有缺损问题，请向购买书店调换；若书店售缺，请与本社发行部联系。联系电话：（010）68279077。质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

Preface

This book has been written for an introductory course in probability and statistics for students of engineering and the physical sciences. Each chapter is keynoted by an introductory statement and has a checklist of key terms (with page numbers) at the end. Important formulas, theorems, and rules are set out in boxes. Confidence intervals are stressed as the major procedure for making inferences. Their properties are carefully described and their interpretation is frequently discussed. The steps for hypothesis testing are clearly and consistently delineated in each application. The examples include the interpretation and calculation of a P-value.

The text has been tested extensively in courses for university students as well as by in-plant training of engineers. The whole book can be covered in a two-semester or three-quarter course consisting of three lectures a week. The book also makes an excellent basis for a one-semester course where the lecturer can choose topics to emphasize theory or application. The author covers most of the first seven chapters, straight-line regression, and the graphic presentation of factorial designs in one semester.

In this seventh edition, we have continued to build on the strengths of the previous editions while adding a few more actual data sets and examples of the application of statistics in scientific investigations. The new data sets, like many of those already in the text, arose in the author's consulting activities or in discussions with scientists and engineers about their statistical problems. Data from some companies have been disguised, but they still retain all of the features necessary to illustrate the statistical methods and the reasoning required to make generalizations from data collected in an experiment. A set of statistical guidelines for correctly applying statistical procedures, and avoiding common pitfalls, has been added to each chapter under the title *Do's and Don'ts.*

To give students an early preview of statistics, descriptive statistics are covered in Chapter 2. Chapters 3 through 6 provide a brief, though rigorous, introduction to the theory of statistics, are suitable for an introductory semester (or quarter) course on probability and statistics. Chapters 7, 8, and 9 are the core material on the key concepts and elementary methods of statistical inference. Chapters 11, 12, and 13 comprise an introduction to some of the standard, though more advanced, topics of experimental design and regression.

The mathematical background expected of the reader is a year course in calculus; actually, calculus is required mainly for Chapter 5 dealing with basic distribution theory in the continuous case.

We wish to thank MINITAB (State College, Pennsylvania) for permission to include commands and output from their MINITAB software program, and

the SAS institute (Gary, North Carolina) for permission to include output from their SAS program.

Thanks go to Paul M. Berthouex, Michael Brajac, Smiley Cheng, Don Ermer, Jim Evans, Cherilyn Hatfield, David Steinberg, and Steve Verrill for contributing data sets in earlier editions. Kuo-Tsung Wu provided valuable help for improving the presentation. Thanks also to Song Yong Sim and Erik Johnson for their help. The current edition benefited from suggestions from George Lobell, Prentice Hall, and the input of the reviewers: Bradford R. Crain, Portland State University; James G. Surles, Texas Tech University; Ulric Lund, California Polytechnic State University; Peter Macdonald, McMaster University; and Jeffrey Proehl. Thanks also go to Ting-Li Lin, who helped integrate some of new MINITAB 14 output into the manuscript and to Moo Chung who helped with some software issues. I am most grateful to Ruojia Li for her considerable help, which enabled this revision to be completed on time.

All revisions in this edition were the responsibility of R. A. Johnson.

Richard A. Johnson
rich@stat.wisc.edu

Contents

Chapter 1

Introduction

CHAPTER OUTLINE

Everything dealing with the collection, processing, analysis, and interpretation of numerical data belongs to the domain of statistics. In engineering, this includes such diversified tasks as calculating the average length of the downtimes of a computer, collecting and presenting data on the numbers of persons attending seminars on solar energy, evaluating the effectiveness of commercial products, predicting the reliability of a rocket, or studying the vibrations of airplane wings.

In Sections 1.2, 1.3 and 1.4 we discuss the recent growth of statistics and, in particular, its applications to problems of engineering. Statistics plays a major role in the improvement of quality of any product or service. An engineer using the techniques described in this book can become much more effective in all phases of work relating to research, development, or production. In Section 1.5 we begin our introduction to statistical concepts by emphasizing the distinction between a population and a sample.

1.1 Why Study Statistics?

Answers provided by statistical analysis can provide the basis for making decisions or choosing actions. For example, city officials might want to know whether the level of lead in the water supply is within safety standards. Because not all of the water can be checked, answers must be based on the partial information from samples of water that are collected for this purpose. As another example, a civil engineer must determine the strength of supports for generators at a power plant. A number of those available must be loaded to failure, and their strengths will provide the basis for assessing the strength of other supports. The proportion of all supports available with strengths that lie below a design limit needs to be determined.

When information is sought, statistical ideas suggest a typical collection process with four crucial steps.

1. **Set clearly defined goals for the investigation.**
2. **Make a plan of what data to collect and how to collect it.**
3. **Apply appropriate statistical methods to extract information from the data.**
4. **Interpret the information and draw conclusions.**

These indispensable steps will provide a frame of reference throughout as we develop the key ideas of statistics. Statistical reasoning and methods can help you become efficient at obtaining information and making useful conclusions.

1.2 Modern Statistics

The origin of statistics can be traced to two areas of interest that, on the surface, have little in common: games of chance and what is now called political science. Mid-eighteenth-century studies in probability, motivated largely by interest in games of chance, led to the mathematical treatment of errors of measurement and the theory that now forms the foundation of statistics. In the same century, interest in the numerical description of political units (cities, provinces, countries, etc.) led to what is now called **descriptive statistics**. At first, descriptive statistics consisted merely of the presentation of data in tables and charts; nowadays, it includes the summarization of data by means of numerical descriptions and graphs.

In recent decades, the growth of statistics has made itself felt in almost every major phase of activity, and the most important feature of its growth has been the shift in emphasis from descriptive statistics to **statistical inference**. Statistical inference concerns generalization based on sample data; it applies to such problems as estimating an engine's average emission of pollutants from trial runs, testing a manufacturer's claim on the basis of measurements performed on samples of his product, and predicting the fidelity of an audio system on the basis of sample data pertaining to the performance of its components.

When one makes a statistical inference, namely, an inference that goes beyond the information contained in a set of data, one must always proceed with caution. One must decide carefully how far to go in generalizing from a given set of data, whether such generalizations are at all reasonable or justifiable, whether it might be wise to wait until there are more data, and so forth. Indeed, some of the most important problems of statistical inference concern the appraisal of the risks and the consequences to which one might be exposed by making generalizations from sample data. This includes an appraisal of the probabilities of making wrong decisions, the chances of making incorrect predictions, and the possibility of obtaining estimates that do not lie within permissible limits.

We shall approach the subject of statistics as a science, developing each statistical idea insofar as possible from its probabilistic foundation, and applying each idea to problems of physical or engineering science as soon as it has been developed. The great majority of the methods we shall use in stating and solving these problems belong to the **classical approach**, because they do not formally take into account the various subjective factors mentioned above. However, we shall endeavor continually to make the reader aware that the subjective factors do exist, and to indicate whenever possible what role they might play in making a final decision. This "bread-and-butter" approach to statistics presents the subject in the form in which it has successfully contributed to engineering science, as well as to the natural and social sciences, in the last half of the twentieth century and beyond.

1.3 Statistics and Engineering

There are few areas where the impact of the recent growth of statistics has been felt more strongly than in engineering and industrial management. Indeed, it would be difficult to overestimate the contributions statistics has made to solving production problems, to the effective use of materials and labor, to basic research, and to the development of new products. As in other sciences, statistics has become a vital tool to engineers. It enables them to understand phenomena subject to variation and to effectively predict or control them.

In this text, our attention will be directed largely toward engineering applications, but we shall not hesitate to refer also to other areas to impress upon the reader the great generality of most statistical techniques: The statistical method used to estimate the average coefficient of thermal expansion of a metal serves also to estimate the average time it takes a secretary to perform a given task, the average thickness of a pelican eggshell, or the average IQ of first-year college students. Similarly, the statistical method used to compare the strength of two alloys serves also to compare the effectiveness of two teaching methods, or the merits of two insect sprays.

1.4 A Case Study: Visually Inspecting Data to Improve Product Quality

This study[1] dramatically illustrates the important advantages gained by appropriately plotting and then monitoring manufacturing data. It concerns a ceramic

[1] Courtesy of Don Ermer

part used in popular coffee makers. This ceramic part is made by filling the cavity between two dies of a pressing machine with a mixture of clay, water, and oil. After pressing, but before the part is dried to a hardened state, critical dimensions are measured. The depth of the slot is of interest here.

Because of natural uncontrolled variation in the clay-water-oil mixture, the condition of the press, differences in operators, and so on, we cannot expect all of the slot measurements to be exactly the same. Some variation in the depth of slots is inevitable, but the depth needs to be controlled within certain limits for the part to fit when assembled.

Slot depth was measured on three ceramic parts selected from production every half hour during the first shift from 6 A.M. to 3 P.M. The data in Table 1.1 were obtained on a Friday. The sample mean, or average, for the first sample of 214, 211, and 218 (thousandths of an inch) is

$$\frac{214+211+218}{3} = \frac{643}{3} = 214.3.$$

The graphical procedure, called an **X-bar** chart, consists of plotting the sample averages versus time order. This plot will indicate when changes have occurred and actions need to be taken to correct the process.

From a prior statistical study, it was known that the process was stable about a value of 217.5 thousandths of an inch. This value will be taken as the central

Table 1.1 Slot depth (thousandths of an inch)

Time	**6:30**	**7:00**	**7:30**	**8:00**	**8:30**	**9:00**	**9:30**	**10:00**
1	214	218	218	216	217	218	218	219
2	211	217	218	218	220	219	217	219
3	218	219	217	219	221	216	217	218
Sum	643	654	653	653	658	653	652	656
$\bar{x}$	214.3	218.0	217.7	217.7	219.3	217.7	217.3	218.7

Time	**10:30**	**11:00**	**11:30**	**12:30**	**1:00**	**1:30**	**2:00**	**2:30**
1	216	216	218	219	217	219	217	215
2	219	218	219	220	220	219	220	215
3	218	217	220	221	216	220	218	214
Sum	653	651	657	660	653	658	655	644
$\bar{x}$	217.7	217.0	219.0	220.0	217.7	219.3	218.3	214.7

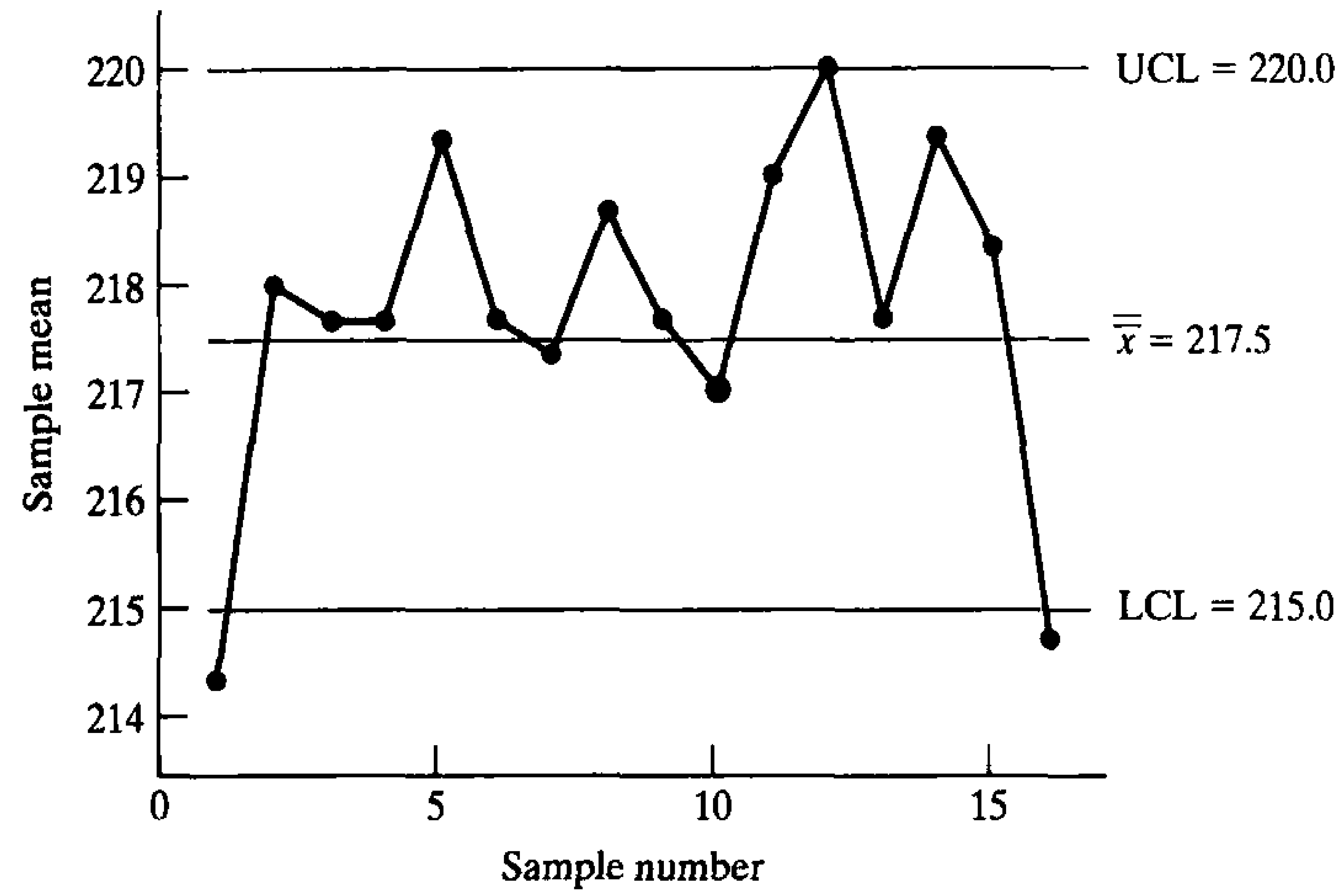


Figure 1.1
X-bar chart for depth

line of the X-bar chart in Figure 1.1.

$$\text{central line}: \ \bar{\bar{x}} = 217.5$$

It was further established that the process was capable of making mostly good ceramic parts if the average slot dimension for a sample remained between certain control limits.

$$\text{Lower control limit: LCL} = 215.0$$

$$\text{Upper control limit: UCL} = 220.0$$

What does the chart tell us? The mean of 214.3 for the first sample, taken at approximately 6:30 A.M., is outside the lower control limit. Further, a measure of the variation in this sample

$$\text{range} = \text{largest} - \text{smallest} = 218 - 211 = 7$$

is large compared to the others. This evidence suggests that the pressing machine had not yet reached a steady state. The control chart suggests that it is necessary to warm up the pressing machine before the first shift begins at 6 A.M. Management and engineering implemented an early start-up and thereby improved the process. The operator and foreman did not have the authority to make this change. Deming claimed that 85% or more of our quality problems are in the system and that the operator and others responsible for the day-to-day operation are responsible for 15% or less of our quality problems.

The X-bar chart further shows that, throughout the day, the process was stable but a little on the high side, although no points were out of control until the last sample of the day. Here an unfortunate oversight occurred. The operator did not report the out-of-control value to either the set-up person or the foreman because it was near the end of her shift and the start of her weekend. She also knew the set-up person was already cleaning up for the end of the shift and that the foreman was likely thinking about going across the street to the Legion Bar for some refreshments as soon as the shift ended. She did not want to ruin anyone's plans so she kept quiet.

On Monday morning when the operator started up the pressing machine, one of the dies broke. The cost of the die was over a thousand dollars. But this was not the biggest cost. When a customer was called and told there would be a delay in delivering the ceramic parts, he canceled the order. Certainly the loss of a customer is an expensive item. Deming referred to this type of cost as the unknown and unknowable, but at the same time it is probably the most important cost of poor quality.

On Friday the chart had predicted a problem. Afterward it was determined that the most likely difficulty was that the clay had dried and stuck to the die, leading to the break. The chart indicated the problem but someone had to act; for a statistical charting procedure to be truly effective action must be taken.

1.5 Two Basic Concepts—Population and Sample

The examples above, where the evaluation of actual information is essential for acquiring new knowledge, motivate the development of statistical reasoning and tools taught in this text. Most experiments and investigations conducted by engineers in the course of investigating, be it a physical phenomenon, production process, or manufactured unit, share some common characteristics.

A first step in any study is to develop a clear, well-defined **statement of purpose**. For example, a mechanical engineer wants to determine whether a new additive will increase the tensile strength of plastic parts produced on an injection molding machine. Not only must the additive increase the tensile strength, it needs to increase it by enough to be of engineering importance. He therefore created the following statement.

Purpose: Determine whether a particular amount of an additive can be found that will increase the tensile strength of the plastic parts by at least 10 pounds per square inch.

In any statement of purpose, try to avoid words like soft, hard, large enough, and so on, which are difficult to quantify. The statement of purpose can help us to decide on what data to collect. For example, the mechanical engineer tried two different amounts of additive and produced 25 specimens of the plastic part with each mixture. The tensile strength was obtained for each of 50 specimens.

Relevant data must be collected. But it is often physically impossible or infeasible from a practical standpoint to obtain a complete set of data. When data are obtained from laboratory experiments, no matter how much experimentation has been performed, more could always be done. To collect an exhaustive set of data related to the damage sustained by all cars of a particular model under collision at a specified speed, every car of that model coming off the production lines would have to be subjected to a collision! In most situations, we must work with only partial information. The distinction between the data actually acquired and the vast collection of all potential observations is a key to understanding statistics.

The source of each measurement is called a **unit**. It is usually an object or a person. To emphasize the term "population" for the entire collection of units, we call the entire collection the **population of units**.

Units and population of units

> **unit**: A single entity, usually an object or person, whose characteristics are of interest.
> **population of units**: The complete collection of units about which information is sought.

Guided by the statement of purpose, we have a **characteristic of interest** for each unit in the population. The characteristic, which could be a qualitative trait, is called a **variable** if it can be expressed as a number.

There can be several characteristics of interest for a given population of units. Some examples are given in Table 1.2.

Table 1.2 Examples of populations, units, and variables

Population	Unit	Variables/Characteristics
All students currently enrolled in school	student	GPA number of credits hours of work per week major right/left-handed
All printed circuit boards manufactured during a month	board	type of defects number of defects location of defects
All campus fast food restaurants	restaurant	number of employees seating capacity hiring/not hiring
All books in library	book	replacement cost frequency of checkout repairs needed

For any population there is the value, for each unit, of a characteristic or variable of interest. For a given variable or characteristic of interest, we call the collection of values, evaluated for every unit in the population, the **statistical population** or just the **population**. This collection of values is the population we will address in all later chapters. Here we refer to the collection of units as the **population of units** when there is a need to differentiate it from the collection of values.

Statistical population

> A **statistical population** is the set of all measurements (or record of some quality trait) corresponding to each unit in the entire population of units about which information is sought.

Generally, any statistical approach to learning about the population begins by taking a sample.