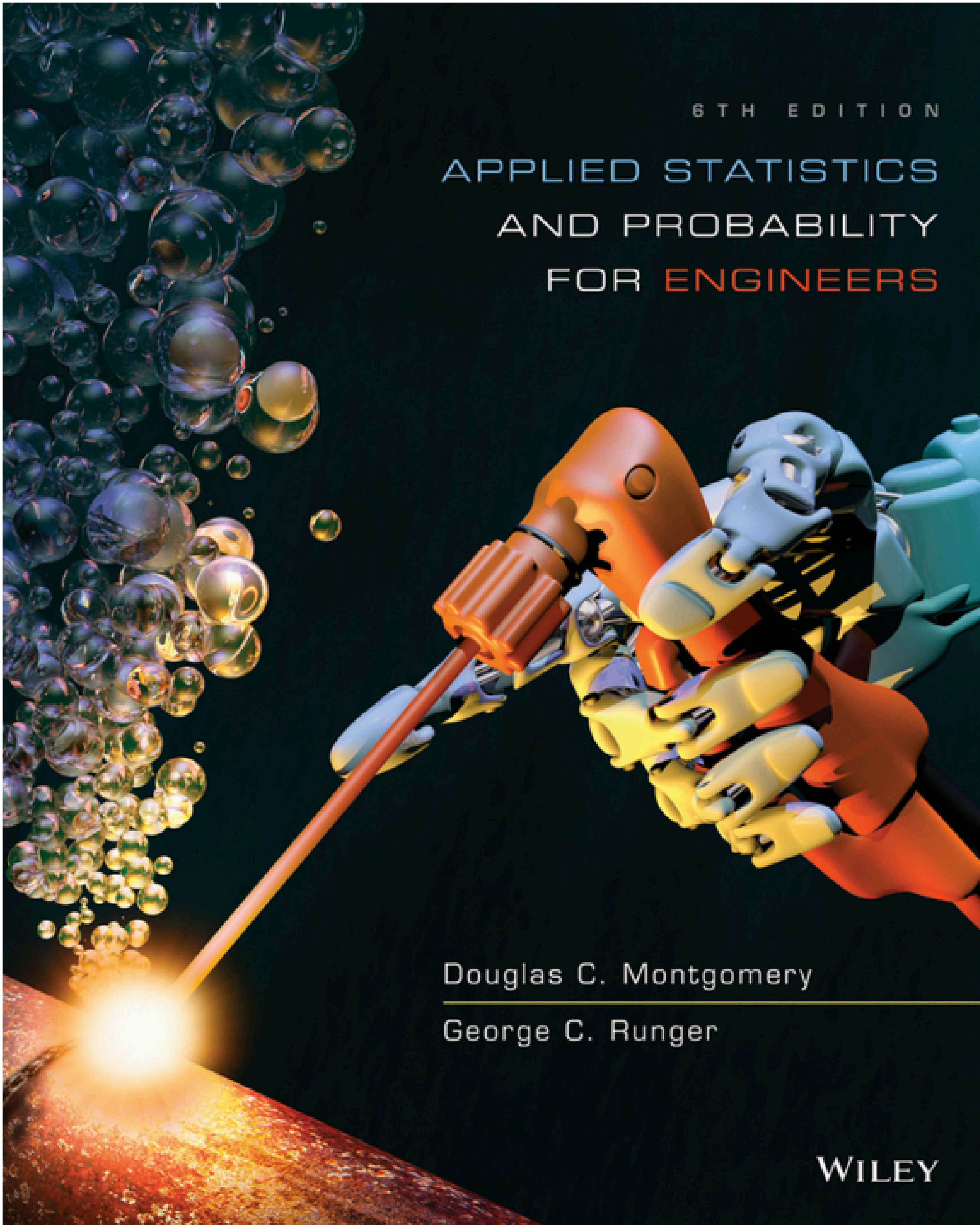
6TH EDITION

APPLIED STATISTICS AND PROBABILITY FOR ENGINEERS

Douglas C. Montgomery

George C. Runger



WILEY

BIOLOGICAL

CHEMICAL

CHEMICAL ENGINEERING

CIVIL ENGINEERING

COMMUNICATIONS, COMPUTERS, AND NETWORKS

ELECTRONICS

ENERGY

ENVIRONMENTAL

MATERIALS

MECHANICAL

(*Text continued at the back of book.*)

Applied Statistics and Probability for Engineers

Sixth Edition

Douglas C. Montgomery

Arizona State University

George C. Runger

Arizona State University

To:

Meredith, Neil, Colin, and Cheryl

Rebecca, Elisa, George and Taylor

VICE PRESIDENT AND PUBLISHER Don Fowley
AQUISITIONS EDITOR Linda Ratts
OPERATIONS MANAGER Melissa Edwards
CONTENT EDITOR Wendy Ashenberg
SENIOR CONTENT MANAGER Kevin Holm
SENIOR PRODUCTION EDITOR Jill Spikereit
EXECUTIVE MARKETING MANAGER Christopher Ruel
PRODUCT DESIGNER Jennifer Welter
DESIGN DIRECTOR Harry Nolan
DESIGNER Kristine Carney
PHOTO EDITOR Felicia Ruocco
PRODUCTION MANAGEMENT SERVICES Laserwords
COVER DESIGN Wendy Lai

Cover Photo © PaulFleet/iStockphoto

Chapter Opener Photos: © babyblueut/iStockphoto, Erik Isakson/Getty Images, Thinkstock Images/Comstock Images/Getty Images, © amriphoto/ iStockphoto, © Ralph125/iStockphoto, © Caro/Alamy, © EasyBuy4u/iStockphoto, © jkullander/iStockphoto, © Alain Nogues/Sygma/Corbis, © HKPNC/iStockphoto, © EdStock/iStockphoto, © skynesher/iStockphoto, © Vasko/iStockphoto, © RedHelga/iStockphoto, © doc-stock/©Corbis, Courtesy of Mike Johnson, www.redbead.com

This book was set by Laserwords Private Limited and printed and bound by RR Donnelley. The cover was printed by RR

Donnelley. This book is printed on acid free paper.

Library of Congress Cataloging-in-Publication Data

ISBN-13 9781118539712
ISBN (BRV)-9781118645062

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Wiley Books by These Authors

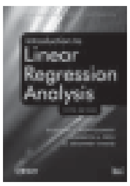
Website: www.wiley.com/college/montgomery

Engineering Statistics, Fifth Edition
by Montgomery and Runger
Introduction to engineering statistics, with topical coverage appropriate for a one-semester course. A modest mathematical level, and an applied approach.

Applied Statistics and Probability for Engineers, Sixth Edition
by Montgomery and Runger
Introduction to engineering statistics, with topical coverage appropriate for eityher a one or two semester course. An applied approach to solving real-world engineering problems.

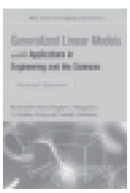
Introduction to Statistical Quality Control, Seventh Edition
by Douglas C. Montgomery
For a first course in statistical quality control. A comprehensive treatment of statistical methodology for quality control and improvement. Some aspects of quality management are also included such as the six-sigma approach.

Design and Analysis of Experiments, Eighth Edition
by Douglas C. Montgomery
An introduction to design and analysis of expieriments, with the modes prerequisite of a first course in statistical methods. For senior and graduate students for a practitioners, to design and analyze experiments for improving the quality and efficiency of working systems.



Introduction to Linear Regression Analysis, Fifth Edition
by Mongomery, Peck and Vining
A comprehensive and thoroughly up to date look at regression analysis still most widely used technique in statistics today.

Response Surface Methodology: Process and Product Optimization Using Designed Experiments, Third Edition
Website: www.wiley.com/college/myers
The exploration and optimization of response surfaces, for graduate courses in experimental design and for applied statisticians, engineers and chemical and physical scientists.



Generalized Linear Models: with Applications in Engineering and the Sciences, Second Edition
by Myers, Montgomery and Vining
Website: www.wiley.com/college/myers
An introductory text or reference on generalized linewar models (GLMs). The range of theoretical topics and applications appeals both to students and practicing professionals.

Introduction to Time Series Analysis and Forecasting

by Montgomery, Jennings and Kulahci
Methods for modeling and analyzing times series data, to draw inferences about the datat and generate forecasts useful to he decision maker. Minitab and SAS are used to illustrate how the methods are implemented in practice. For Advanced undergrad/first-year graduate, with a prereq uisite of basic statistical methods. Portions of the book require calculus and matrix algebra.

INTENDED AUDIENCE

This is an introductory textbook for a first course in applied statistics and probability for undergraduate students in engineering and the physical or chemical sciences. These indi viduals play a significant role in designing and developing new products and manufacturing systems and processes, and they also improve existing systems. Statistical methods are an important tool in these activities because they provide the engineer with both descriptive and analytical methods for dealing with the variability in observed data. Although many of the methods we present are fundamental to statistical analysis in other disciplines, such as business and management, the life sciences, and the social sciences, we have elected to focus on an engineering-oriented audience. We believe that this approach will best serve students in engineering and the chemical/physical sciences and will allow them to concentrate on the many applications of statistics in these disciplines. We have worked hard to ensure that our examples and exercises are engineering- and science-based, and in almost all cases we have used examples of real data—either taken from a published source or based on our consulting experiences.

We believe that engineers in all disciplines should take at least one course in statistics. Unfortunately, because of other requirements, most engineers will only take one statistics course. This book can be used for a single course, although we have provided enough material for two courses in the hope that more students will see the important applications of statistics in their everyday work and elect a second course. We believe that this book will also serve as a useful reference.

We have retained the relatively modest mathematical level of the first five editions. We have found that engineering students who have completed one or two semesters of calculus and have some knowledge of matrix algebra should have no difficulty reading all of the text. It is our intent to give the reader an understanding of the methodology and how to apply it, not the mathematical theory. We have made many enhancements in this edition, including

reorganiz
ing and rewriting major portions of the book and adding a number of new exercises.

ORGANIZATION OF THE BOOK

Perhaps the most common criticism of engineering statistics texts is that they are too long. Both instructors and students complain that it is impossible to cover all of the topics in the book in one or even two terms. For authors, this is a serious issue because there is great variety in both the content and level of these courses, and the decisions about what material to delete without limiting the value of the text are not easy. Decisions about which topics to include in this edition were made based on a survey of instructors.

Chapter 1 is an introduction to the field of statistics and how engineers use statistical meth odology as part of the engineering problem-solving process. This chapter also introduces the reader to some engineering applications of statistics, including building empirical models, designing engineering experiments, and monitoring manufacturing processes. These topics are discussed in more depth in subsequent chapters.

Chapters 2, 3, 4, and 5 cover the basic concepts of probability, discrete and continuous random variables, probability distributions, expected values, joint probability distributions, and independence. We have given a reasonably complete treatment of these topics but have avoided many of the mathematical or more theoretical details.

Chapter 6 begins the treatment of statistical methods with random sampling; data sum mary and description techniques, including stem-and-leaf plots, histograms, box plots, and probability plotting; and several types of time series plots. Chapter 7 discusses sampling dis tributions, the central limit theorem, and point estimation of parameters. This chapter also introduces some of the important properties of estimators, the method of maximum likeli hood, the method of moments, and Bayesian estimation.

Chapter 8 discusses interval estimation for a single sample. Topics included are confidence intervals for means, variances or standard deviations, proportions, prediction intervals, and tol erance intervals. Chapter 9 discusses hypothesis tests for a single sample. Chapter 10 presents tests and confidence intervals for two samples. This material has been extensively rewritten and reorganized. There is detailed information and examples of methods for determining appropri ate sample sizes. We want the student to become familiar with how these techniques are used to solve real-world engineering problems and to get some understanding of the concepts behind them. We give a logical, heuristic development of the procedures rather than a formal, mathe matical one. We have also included some material on nonparametric methods in these chapters.

Chapters 11 and 12 present simple and multiple linear regression including model ade quacy checking and regression model diagnostics and an introduction to logistic regression. We use matrix algebra throughout the multiple regression material (Chapter 12) because it is the only easy way to understand the concepts presented. Scalar arithmetic presentations of multiple regression are awkward at best, and we have found that undergraduate engineers are exposed to enough matrix algebra to understand the presentation of this material.

Chapters 13 and 14 deal with single- and multifactor experiments, respectively. The notions of randomization, blocking, factorial designs, interactions, graphical data analysis, and frac tional factorials are emphasized. Chapter 15 introduces statistical quality control, emphasiz ing the control chart and the fundamentals of statistical process control.

WHAT'S NEW IN THIS EDITION

We received much feedback from users of the fifth edition of the book, and in response we have made substantial changes in this new edition.

have added material on the bootstrap and its use in constructing confidence intervals.

P-value in hypothesis testing. Many sections of several chapters were rewritten to reflect this.

try to make the concepts easier to understand.

ing, a technique widely used in the biopharmaceutical industry, but which has widespread applications in other areas.

P-values when performing mutiple tests is incuded.

tions of the results.

FEATURED IN THIS BOOK

Definitions, Key Concepts, and

Throughout the text, definitions a
concepts and equations are highli
box to emphasize their importanc

Learning Objectives

Learning Objectives at the start of each chapter guide the students in what they are expected to take away from this chapter and serve as a study reference.

Seven-Step Procedure for Hypothesis Testing

The text introduces a sequence of seven steps in applying hypothesis-testing methodology and explicitly exhibits this procedure in examples.

Figures

Numerous figures throughout

the text illustrate statistical concepts in multiple formats.

Computer Output
Example throughout the book, use computer output to illustrate the role of modern statistical software.

Example Problems
A set of example problems provides the student with detailed solutions and comments for interesting, real-world situations. Brief practical interpretations have been added in this edition.

Exercises
Each chapter has an extensive collection of exercises, including **end-of-section exercises** that emphasize the material in that section, **supplemental exercises** at the end of the chapter that cover the scope of chapter topics and require the student to make a decision about the approach they will use to solve the problem, and **mind-expanding exercises** that often require the student to extend the text material somewhat or to apply it in a novel situation. Answers are provided to most odd numbered exercises in Appendix C in the text, and the *WileyPLUS* online learning environment includes for students complete detailed solutions to selected exercises.

Important Terms and Concepts
At the end of each chapter is a list of important terms and concepts for an easy self-check and study tool.

STUDENT RESOURCES

the book Web site at www.wiley.com/college/montgomery to access these materials.

Student Solutions Manual may be purchased from the Web site at www.wiley.com/college/ montgomery.

INSTRUCTOR RESOURCES

The following resources are available only to instructors who adopt the text: **Solutions Manual** All solutions to the exercises in the text.

Data Sets Data sets for all examples and exercises in the text.

Image Gallery of **Text Figures**

PowerPoint Lecture Slides

Section on Logistic Regression

These instructor-only resources are password-protected. Visit the instructor section of the book Web site at www.wiley.com/college/montgomery to register for a password to access these materials.

COMPUTER SOFTWARE

We have used several different packages, including Excel, to demonstrate computer usage. Minitab can be used for most exercises. A student version of Minitab is available as an option to purchase in a set with this text. Student versions of software often do not have all the functionality that full versions do. Consequently, student versions may not support all the concepts presented in this text. If you would like to adopt for your course the set of this text with the student version of Minitab, please contact your local Wiley representative at www.wiley.com/college/rep.

Alternatively, students may find information about how to purchase the professional version of the software for academic use at www.minitab.com.

WileyPLUS

This online teaching and learning environment integrates the **entire digital textbook** with the most effective instructor and student resources to fit every learning style. With *WileyPLUS*:

ments, grade tracking, and more.

WileyPLUS can complement your current textbook or replace the printed text altogether.

For Students

Personalize the learning experience

Different learning styles, different levels of proficiency, different levels of preparation—each of your students is unique. *WileyPLUS* empowers them to take advantage of their individual strengths:

immediate feedback and remediation when needed.

erences and encourage more active learning.

WileyPLUS includes many opportunities for self-assessment linked to the relevant portions of the text. Students can take control of their own learning and practice until they master the material.

For Instructors

Personalize the teaching experience
WileyPLUS empowers you with the tools and resources you need to make your teaching even more effective:

als to your *WileyPLUS* course.

WileyPLUS you can identify those students who are falling behind and intervene accordingly, without having to wait for them to come to office hours.

WileyPLUS simplifies and automates such tasks as student performance assessment, mak ing assignments, scoring student work, keeping grades, and more.

COURSE SYLLABUS SUGGESTIONS

course on statistics for engineers vary widely, as do the abilities of different groups of stu dents. Therefore, we hesitate to give too much advice, but will explain how we use the book.

We believe that a first course in statistics for engineers should be primarily an applied statistics course, not a probability course. In our one-semester course we cover all of Chapter 1 (in one or two lectures); overview the material on probability, putting most of the emphasis on the normal distribution (six to eight lectures); discuss most of Chapters 6 through 10 on confidence intervals and tests (twelve to fourteen lectures); introduce regression models in Chapter 11 (four lectures); give an introduction to the design of experiments from Chapters 13 and 14 (six lectures); and present the basic concepts of statistical process control, including the Shewhart control chart from Chapter 15 (four lectures). This leaves about three to four periods for exams and review. Let us emphasize that the purpose of this course is to introduce engineers to how statistics can be used to solve real-world engineering problems, not to weed out the less mathematically gifted students. This course is not the “baby math-stat” course that is all too often given to engineers.

If a second semester is available, it is possible to cover the entire book, including much of the supplemental material, if appropriate for the audience. It would also be possible to assign and work many of the homework problems in class to reinforce the understanding of the con cepts. Obviously, multiple regression and more design of experiments would be major topics in a second course.

USING THE COMPUTER

In practice, engineers use computers to apply statistical methods to solve problems. Therefore, we strongly recommend that the computer be integrated into the class. Throughout the book we have presented typical example of the output that can be obtained with modern statistical software. In teaching, we have used a variety of software packages, including Minitab, Stat graphics, JMP, and Statistica. We did not clutter up the book with operational details of these different packages because how the instructor integrates the software into the class is ultimate ly more important than which package is used. All text data are available in electronic form on the textbook Web site. In some chapters, there are problems that we feel should be worked using computer software. We have marked these problems with a special icon in the margin.

In our own classrooms, we use the computer in almost every lecture and demonstrate how the technique is implemented in software as soon as it is discussed in the lecture. Student versions of many statistical software packages are available at low cost, and students can either purchase their own copy or use the products available through the institution. We have found that this greatly improves the pace of the course and student understanding of the material.

Users should be aware that final answers may differ slightly due to different numerical preci sion and rounding protocols among softwares.

ACKNOWLEDGMENTS

We would like to express our grateful appreciation to the many organizations and individuals who have contributed to this book. Many instructors who used the previous editions provided excellent suggestions that we have tried to incorporate in this revision.

We would like to thank the following who assisted in contributing to and/or reviewing material for the *WileyPLUS* course:

Michael DeVasher, *Rose-Hulman Institute of Technology*
Craig Downing, *Rose-Hulman Institute of Technology*
Julie Fortune, *University of Alabama in Huntsville*
Rubin Wei, *Texas A&M University*

We would also like to thank the following for their assistance in checking the accuracy and completeness of the exercises and the solutions to exercises.

Dr. Abdelaziz Berrado
Dr. Connie Borror
Aysegul Demirtas
Kerem Demirtas
Patrick Egbunonu, Sindhura Gangu
James C. Ford
Dr. Alejandro Heredia-Langner
Dr. Jing Hu
Dr. Busaba Laungrungrong
Dr. Fang Li
Dr. Nuttha Lurponglukana

Dr. Lora Zimmer

We are also indebted to Dr. Smiley Cheng for permission to adapt many of the statistical tables from his excellent book (with Dr. James Fu), *Statistical Tables for Classroom and Exam Room*. Wiley, Prentice Hall, the Institute of Mathematical Statistics, and the editors of Biomet rics allowed us to use copyrighted material, for which we are grateful.

Douglas C. Montgomery

George C. Runger

Contents

Contents

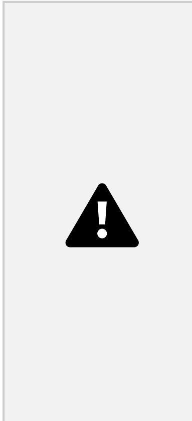
Chapter 12 Multiple Linear Regression 477

Contents

Chapter 13 Design and Analysis of Single-Factor Experiments: The Analysis of Variance 539

Chapter 14 Design of Experiments with Several Factors 575

Chapter 15 Statistical Quality Control 663



Chapter Outline

1

The Role of Statistics in Engineering

Time **1-3 Mechanistic and Empirical Models 1-4 Probability and Probability Models**

Statistics is a science that helps us make decisions and draw conclusions in the presence of variability. For example, civil engineers working in the transportation fi eld are concerned

1-1 The Engineering Method and Statistical Thinking

1-2 Collecting Engineering Data

1-2.1 Basic Principles
1-2.2 Retrospective Study
1-2.3 Observational Study
1-2.4 Designed Experiments
1-2.5 Observing Processes Over

about the capacity of regional highway systems. A typical problem related to transportation would involve data regarding this specifi c system's number of nonwork, home-based trips, the number of persons per household, and the number of vehi

cles per household. The objective would be to produce a trip generation model relating trips to the number of persons per household and the number of vehicles per household. A statis tical technique called *regression analysis* can be used to con struct this model. The trip-generation model is an important tool for transportation systems planning. Regression methods are among the most widely used statistical techniques in engi

neering. They are presented in Chapters 11 and 12. The hospital emergency department (ED) is an important part of the healthcare delivery system. The process by which patients arrive at the ED is highly variable and can depend on the hour of the day and the day of the week, as well as on longer-term cyclical variations. The service process is also highly variable, depending on the types of services that the patients require, the number of patients in the ED, and how the ED is staffed and organized. An ED's capacity is also limited; consequently, some patients experience long waiting times. How long do patients wait, on average? This is an important question for healthcare providers. If waiting times become excessive, some patients will leave without receiving treatment LWOT. Patients who LWOT are a serious problem, because they do not have their medical concerns addressed and are at risk for further problems and complications. Therefore, another

important question is: What proportion of patients LWOT from the ED? These questions can be solved by employing probability models to describe the ED, and from these models very precise estimates of waiting times and the number of patients who LWOT can be obtained. Probability models that can be used to solve these types of problems are discussed in Chapters 2 through 5.

The concepts of probability and statistics are powerful ones and contribute extensively to the solutions of many types of engineering problems. You will encounter many examples of these applications in this book.

Learning Objectives

After careful study of this chapter, you should be able to do the following:

1. Identify the role that statistics can play in the engineering problem-solving process 2. Discuss how variability affects the data collected and used for making engineering decisions 3. Explain the difference between enumerative and analytical studies

4. Discuss the different methods that engineers use to collect data
5. Identify the advantages that designed experiments have in comparison to other methods of collecting engineering data
6. Explain the differences between mechanistic models and empirical models
7. Discuss how probability and probability models are used in engineering and science

1-1 The Engineering Method and Statistical Thinking

An engineer is someone who solves problems of interest to society by the efficient application of scientific principles. Engineers accomplish this by either refining an existing product or process or by designing a new product or process that meets customers' needs. The **engineering**, or **scientific**, **method** is the approach to formulating and solving these problems. The steps in the engineering method are as follows:

1. Develop a clear and concise description of the problem.

2. Identify, at least tentatively, the important factors that affect this problem or that may play a role in its solution.

3. Propose a model for the problem, using scientific or engineering knowledge of the phenomenon being studied. State any limitations or assumptions of the model.

4. Conduct appropriate experiments and collect data to test or validate the tentative model or conclusions made in steps 2 and 3.

5. Refine the model on the basis of the observed data.

6. Manipulate the model to assist in developing a solution to the problem.

7. Conduct an appropriate experiment to confirm that the proposed solution to the problem is both effective and efficient.

8. Draw conclusions or make recommendations based on the problem solution.

The steps in the engineering method are shown in Fig. 1-1. Many engineering sciences employ the engineering method: the mechanical sciences (statics, dynamics), fluid science, thermal science, electrical science, and the science of materials. Notice that the engineer ing method features a strong interplay among the problem, the factors that may influence its solution, a model of the phenomenon, and experimentation to verify the adequacy of the model and the proposed solution to the problem. Steps 2–4 in Fig. 1-1 are enclosed in a box, indicating that several cycles or iterations of these steps may be required to obtain the final solution. Consequently, engineers must know how to efficiently plan experiments, collect data, analyze and interpret the data, and understand how the observed data relate to the model they have proposed for the problem under study.

description
Identify the important factors
Propose or refine a model
Manipulate the model
Confirm the solution
Conclusions and recommendations
Develop a clear

FIGURE 1-1 The engineering method.

The Science of Data Variability

Conduct experiments

The field of **statistics** deals with the collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes. In simple terms, **statistics is the sci ence of data**. Because many aspects of engineering practice involve working with data, obviously knowledge of statistics is just as important to an engineer as are the other engineering sciences. Specifically, statistical techniques can be powerful aids in designing new products and systems, improving existing designs, and designing, developing, and improving production processes.

Statistical methods are used to help us describe and understand **variability**. By variability, we mean that successive observations of a system or phenomenon do not produce exactly the same result. We all encounter variability in our everyday lives, and **statistical thinking** can give us a useful way to incorporate this variability into our decision-making processes. For example, con sider the gasoline mileage performance of your car. Do you always get exactly the same mileage performance on every tank of fuel? Of course not — in fact, sometimes the mileage performance varies considerably. This observed variability in gasoline mileage depends on many factors, such as the type of driving that has occurred most recently (city versus highway), the changes in the vehicle's condition over time (which could include factors such as tire inflation, engine com pression, or valve wear), the brand and/or octane number of the gasoline used, or possibly even the weather conditions that have been recently experienced. These factors represent potential **sources of variability** in the system. Statistics provides a framework for describing this vari ability and for learning about which potential sources of variability are the most important or which have the greatest impact on the gasoline mileage performance.

We also encounter variability in dealing with engineering problems. For example, suppose that an engineer is designing a nylon connector to be used in an automotive engine application. The engineer is considering establishing the design specification on wall thickness at 3 32 inch but is somewhat uncertain about the effect of this decision on the connector pull-off force. If the pull-off force is too low, the connector may fail when it is installed in an engine. Eight prototype units are produced and their pull-off forces measured, resulting in the following data (in pounds): 12 6 12 9 13 4 12 3 13 6 13 5 12 6 13 1 ., ., ., ., ., ., ., . . As we anticipated, not all of the prototypes have the same pull-off force. We say that there is variability in the pull-off force measurements. Because the pull-off force measurements exhibit variability, we consider the pull-off force to be a **random variable**. A convenient way to think of a random variable, say *X*, that represents a measurement is by using the model

$$X = \mu + \epsilon \quad (1\text{-}1)$$

where μ is a constant and ϵ is a random disturbance. The constant remains the same with every measurement, but small changes in the environment, variance in test equipment, differences in the individual parts themselves, and so forth change the value of ϵ. If there were no distur bances, ϵ would always equal zero and *X* would always be equal to the constant μ. However, this never happens in the real world, so the actual measurements *X* exhibit variability. We often need to describe, quantify, and ultimately reduce variability.

Figure 1-2 presents a **dot diagram** of these data. The dot diagram is a very useful plot for displaying a small body of data—say, up to about 20 observations. This plot allows us to easily see two features of the data: the **location**, or the middle, and the **scatter** or **variability**. When the number of observations is small, it is usually difficult to identify any specific patterns in the variability, although the dot diagram is a convenient way to see any unusual data features.

12 13 14 15 Pull-off force

12 13 14 15 Pull-off force

$= \frac{3}{32}$ inch

$\frac{1}{8} =$ inch

FIGURE 1-2 Dot diagram of the pull-off force data when wall thickness is 3 32 inch.

FIGURE 1-3 Dot diagram of pull-off force for two wall thicknesses.

Population and Samples

The need for statistical thinking arises often in the solution of engineering problems. Consider the engineer designing the connector. From testing the prototypes, he knows that the average pull off force is 13.0 pounds. However, he thinks that this may be too low for the intended application, so he decides to consider an alternative design with a thicker wall, 1 8 inch in thickness. Eight pro totypes of this design are built, and the observed pull-off force measurements are 12.9, 13.7, 12.8, 13.9, 14.2, 13.2, 13.5, and 13.1. The average is 13.4. Results for both samples are plotted as dot diagrams in Fig. 1-3. This display gives the impression that increasing the wall thickness has led to an increase in pull-off force. However, there are some obvious questions to ask. For instance, how do we know that another sample of prototypes will not give different results? Is a sample of eight prototypes adequate to give reliable results? If we use the test results obtained so far to conclude that increasing the wall thickness increases the strength, what risks are associated with this deci sion? For example, is it possible that the apparent increase in pull-off force observed in the thicker prototypes is due only to the inherent variability in the system and that increasing the thickness of the part (and its cost) really has no effect on the pull-off force?

Often, physical laws (such as Ohm's law and the ideal gas law) are applied to help design prod ucts and processes. We are familiar with this reasoning from general laws to specific cases. But it is also important to reason from a specific set of measurements to more general cases to answer the previous questions. This reasoning comes from a **sample** (such as the eight connectors) to a **population** (such as the connectors that will be in the products that are sold to customers). The reasoning is referred to as **statistical inference**. See Fig. 1-4. Historically, measurements were obtained from a sample of people and generalized to a population, and the terminology has remained. Clearly, reasoning based on measurements from some objects to measurements on all objects can result in errors (called *sampling errors*). However, if the sample is selected properly, these risks can be quantified and an appropriate sample size can be determined.

1-2 Collecting Engineering Data

1-2.1 BASIC PRINCIPLES

In the previous subsection, we illustrated some simple methods for summarizing data. Some times the data are all of the observations in the population. This results in a **census**. However, in the engineering environment, the data are almost always a **sample** that has been selected from the population. Three basic methods of collecting data are

A **retrospective study** using historical data

An **observational study**

A **designed experiment**

inference is one type of reasoning.

designs

Population

Physical laws

Statistical inference

Types of reasoning

Sample

Product

FIGURE 1-4 Statistical

An effective data-collection procedure can greatly simplify the analysis and lead to improved understanding of the population or process that is being studied. We now consider some examples of these data-collection methods.

1-2.2 RETROSPECTIVE STUDY

Montgomery, Peck, and Vining (2012) describe an acetone-butyl alcohol distillation column for which concentration of acetone in the distillate (the output product stream) is an important variable. Factors that may affect the distillate are the reboil temperature, the condensate temperature, and the reflux rate. Production personnel obtain and archive the following records:

The concentration of acetone in an hourly test sample of output product

The reboil temperature log, which is a record of the reboil temperature over time

The condenser temperature controller log

The nominal reflux rate each hour

The reflux rate should be held constant for this process. Consequently, production personnel change this very infrequently.

Hazards of Using Historical Data

A retrospective study would use either all or a sample of the historical process data archived over some period of time. The study objective might be to discover the relationships among the two temperatures and the reflux rate on the acetone concentration in the output product stream. However, this type of study presents some problems:

1. We may not be able to see the relationship between the reflux rate and acetone concentration because the reflux rate did not change much over the historical period.

2. The archived data on the two temperatures (which are recorded almost continuously) do not correspond perfectly to the acetone concentration measurements (which are made hourly). It may not be obvious how to construct an approximate correspondence.

3. Production maintains the two temperatures as closely as possible to desired targets or set points. Because the temperatures change so little, it may be difficult to assess their real impact on acetone concentration.

4. In the narrow ranges within which they do vary, the condensate temperature tends to increase with the reboil temperature. Consequently, the effects of these two process vari ables on acetone concentration may be difficult to separate.

As you can see, a retrospective study may involve a significant amount of **data**, but those data may contain relatively little useful **information** about the problem. Furthermore, some of the relevant data may be missing, there may be transcription or recording errors resulting in **outli ers** (or unusual values), or data on other important factors may not have been collected and archived. In the distillation column, for example, the specific concentrations of butyl alcohol and acetone in the input feed stream are very important factors, but they are not archived because the concentrations are too hard to obtain on a routine basis. As a result of these types of issues, statistical analysis of historical data sometimes identifies interesting phenomena, but solid and reliable explanations of these phenomena are often difficult to obtain.

1-2.3 OBSERVATIONAL STUDY

In an observational study, the engineer observes the process or population, disturbing it as little as possible, and records the quantities of interest. Because these studies are usually conducted for a relatively short time period, sometimes variables that are not routinely measured can be included. In the distillation column, the engineer would design a form to record the two temperatures and the reflux rate when acetone concentration measurements are made. It may even be possible to measure the input feed stream concentrations so that the impact of this factor could be studied.

Generally, an observational study tends to solve problems 1 and 2 and goes a long way toward obtaining accurate and reliable data. However, observational studies may not help resolve problems 3 and 4.

1-2.4 DESIGNED EXPERIMENTS

In a designed experiment, the engineer makes *deliberate* or *purposeful changes* in the controlla ble variables of the system or process, observes the resulting system output data, and then makes an inference or decision about which variables are responsible for the observed changes in output performance. The nylon connector example in Section 1-1 illustrates a **designed experiment**; that is, a deliberate change was made in the connector's wall thickness with the objective of dis covering whether or not a stronger pull-off force could be obtained. Experiments designed with basic principles such as **randomization** are needed to establish **cause-and-effect** relationships. Much of what we know in the engineering and physical-chemical sciences is developed through testing or experimentation. Often engineers work in problem areas in which no scien tific or engineering theory is directly or completely applicable, so experimentation and obser vation of the resulting data constitute the only way that the problem can be solved. Even when there is a good underlying scientific theory that we may rely on to explain the phenomena of interest, it is almost always necessary to conduct tests or experiments to confirm that the the ory is indeed operative in the situation or environment in which it is being applied. Statistical thinking and statistical methods play an important role in planning, conducting, and analyzing the data from engineering experiments. Designed experiments play a very important role in engineering design and development and in the improvement of manufacturing processes.

For example, consider the problem involving the choice of wall thickness for the nylon connec tor. This is a simple illustration of a designed experiment. The engineer chose two wall thicknesses for the connector and performed a series of tests to obtain pull-off force measurements at each wall thickness. In this simple **comparative experiment**, the engineer is interested in determining whether there is any difference between the 3 32- and 1 8-inch designs. An approach that could be used in analyzing the data from this experiment is to compare the mean pull-off force for the 3 32

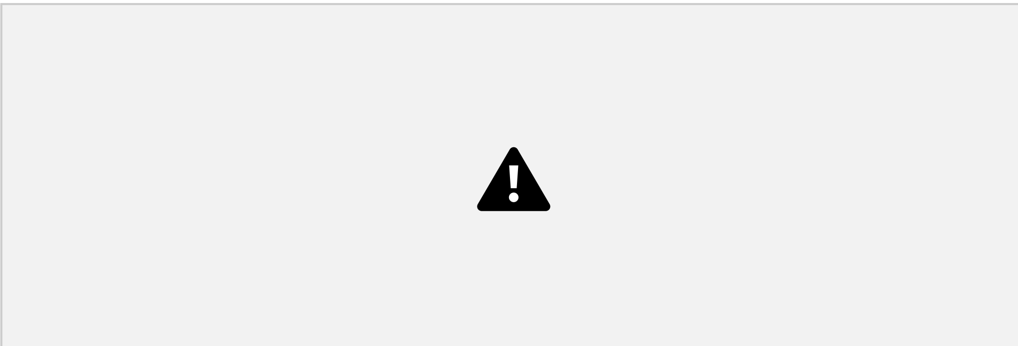
-inch design to the mean pull-off force for the 1 8-inch design using statistical **hypothesis testing**, which is discussed in detail in Chapters 9 and 10. Generally, a **hypothesis** is a statement about some aspect of the system in which we are interested. For example, the engineer might want to know if the mean pull-off force of a 3 32-inch design exceeds the typical maximum load expected to be encountered in this application, say, 12.75 pounds. Thus, we would be interested in testing the hypothesis that the mean strength exceeds 12.75 pounds. This is called a **single-sample hypothesis**

testing problem. Chapter 9 presents techniques for this type of problem. Alternatively, the engineer might be interested in testing the hypothesis that increasing the wall thickness from 3 32 to 1 8 inch results in an increase in mean pull-off force. It is an example of a **two-sample hypothesis-testing problem**. Two-sample hypothesis-testing problems are discussed in Chapter 10.

Designed experiments offer a very powerful approach to studying complex systems, such as the distillation column. This process has three factors—the two temperatures and the reflux rate—and we want to investigate the effect of these three factors on output acetone concentra tion. A good experimental design for this problem must ensure that we can separate the effects of all three factors on the acetone concentration. The specified values of the three factors used in the experiment are called **factor levels.** Typically, we use a small number of levels such as two or three for each factor. For the distillation column problem, suppose that we use two lev els, "high" and "low" (denoted +1 and −1, respectively), for each of the three factors. A very reasonable experiment design strategy uses every possible combination of the factor levels to form a basic experiment with eight different settings for the process. This type of experiment is called a **factorial experiment.** See Table 1-1 for this experimental design.

Figure 1-5 illustrates that this design forms a cube in terms of these high and low levels. With each setting of the process conditions, we allow the column to reach equilibrium, take a sample of the product stream, and determine the acetone concentration. We then can draw

Interaction can be a Key Element in Problem Solving



1-1 The Designed Experiment (Factorial Design) for the Distillation Column

Reboil Temp.	Condensate Temp.	Reflux Rate
–1	–1	–1
+1	–1	–1
–1	+1	–1
+1	+1	–1
–1	–1	+1
+1	–1	+1
–1	+1	+1
+1	+1	+1

specifi c inferences about the effect of these factors. Such an approach allows us to proactively study a population or process.

An important advantage of factorial experiments is that they allow one to detect an **interac tion** between factors. Consider only the two temperature factors in the distillation experiment. Suppose that the response concentration is poor when the reboil temperature is *low,* regardless of the condensate temperature. That is, the condensate temperature has no effect when the reboil temperature is *low*. However, when the reboil temperature is *high,* a *high* condensate tempera ture generates a good response, but a *low* condensate temperature generates a poor response. That is, the condensate temperature changes the response when the reboil temperature is *high*. The effect of condensate temperature depends on the setting of the reboil temperature, and these two factors are said to interact in this case. If the four combinations of *high* and *low* reboil and condensate temperatures were not tested, such an interaction would not be detected.

We can easily extend the factorial strategy to more factors. Suppose that the engineer wants to consider a fourth factor, type of distillation column. There are two types: the standard one and a newer design. Figure 1-6 illustrates how all four factors—reboil temperature, conden sate temperature, refl ux rate, and column design—could be investigated in a factorial design. Because all four factors are still at two levels, the experimental design can still be represented geometrically as a cube (actually, it's a *hypercube*). Notice that as in any factorial design, all possible combinations of the four factors are tested. The experiment requires 16 trials.

Generally, if there are k factors and each has two levels, a factorial experimental design will require 2^k runs. For example, with $k = 4$, the 2^4 design in Fig. 1-6 requires 16 tests. Clearly, as the number of factors increases, the number of trials required in a factorial experiment increases rap idly; for instance, eight factors each at two levels would require 256 trials. This quickly becomes unfeasible from the viewpoint of time and other resources. Fortunately, with four to fi ve or more factors, it is usually unnecessary to test all possible combinations of factor levels. A **fractional factorial experiment** is a variation of the basic factorial arrangement in which only a subset of the factor combinations is actually tested. Figure 1-7 shows a fractional factorial experimental design for the four-factor version of the distillation experiment. The circled test combinations in this fi gure are the only test combinations that need to be run. This experimental design requires only 8 runs instead of the original 16; consequently it would be called a **one-half fraction**. This is an excellent experimental design in which to study all four factors. It will provide good information about the individual effects of the four factors and some information about how these factors interact.

+1 +1

Condensate

Reflux rate

–1 –1

temperature +1

–1

Reboil temperature

FIGURE 1-5 The factorial design for the distillation column.

 Column design

Old New Reflux rate

Condensate temperature 11

FIGURE 1-6 A four-factorial experiment for the distillation column.

21
11
21

21 11

Reboil temperature

Factorial and fractional factorial experiments are used extensively by engineers and scientists in industrial research and development, where new technology, products, and processes are designed and developed and where existing products and processes are improved. Since so much engineer ing work involves testing and experimentation, it is essential that all engineers understand the basic principles of planning efficient and effective experiments. We discuss these principles in Chapter 13. Chapter 14 concentrates on the factorial and fractional factorials that we have introduced here.

1-2.5 Observing Processes Over Time

Often data are collected over time. In this case, it is usually very helpful to plot the data versus time in a **time series plot.** Phenomena that might affect the system or process often become more visible in a time-oriented plot and the concept of stability can be better judged.

Figure 1-8 is a dot diagram of acetone concentration readings taken hourly from the distil lation column described in Section 1-2.2. The large variation displayed on the dot diagram indicates considerable variability in the concentration, but the chart does not help explain the reason for the variation. The time series plot is shown in Fig. 1-9. A shift in the process mean level is visible in the plot and an estimate of the time of the shift can be obtained.

W. Edwards Deming, a very influential industrial statistician, stressed that it is important to understand the nature of variability in processes and systems over time. He conducted an experiment in which he attempted to drop marbles as close as possible to a target on a table. He used a funnel mounted on a ring stand and the marbles were dropped into the funnel. See Fig. 1-10. The funnel was aligned as closely as possible with the center of the target. He then used two different strategies to operate the process. (1) He never moved the funnel. He just dropped one marble after another and recorded the distance from the target. (2) He dropped the first marble and recorded its location relative to the target. He then moved the funnel an equal and opposite distance in an attempt to compensate for the error. He continued to make this type of adjustment after each marble was dropped.

Unnecessary Adjustments Can Increase Variability

After both strategies were completed, he noticed that the variability of the distance from the target for strategy 2 was approximately twice as large than for strategy 1. The adjustments to the funnel increased the deviations from the target. The explanation is that the error (the devia tion of the marble's position from the target) for one marble provides no information about the error that will occur for the next marble. Consequently, adjustments to the funnel do not decrease future errors. Instead, they tend to move the funnel farther from the target.

Column design

FIGURE 1-7 A fractional factorial experiment for the connector wall thickness problem.

Old New

Condensate temperature 11

Reflux rate

21
11
21

21 11

Reboil temperature

Acetone concentration

90

80

80.5 84.0 87.5 91.0 94.5 98.0 x

20

30

10

Acetone concentration

FIGURE 1-9 A time series plot of concentration provides more information than the dot diagram.

FIGURE 1-8 The dot diagram illustrates variation but does not identify the problem.

Observation number (hour)

This interesting experiment points out that adjustments to a process based on random dis turbances can actually *increase* the variation of the process. This is referred to as **overcontrol** or **tampering**. Adjustments should be applied only to compensate for a nonrandom shift in the process—then they can help. A computer simulation can be used to demonstrate the les sons of the funnel experiment. Figure 1-11 displays a time plot of 100 measurements (denoted as y) from a process in which only random disturbances are present. The target value for the process is 10 units. The figure displays the data with and without adjustments that are applied to the process mean in an attempt to produce data closer to target. Each adjustment is equal and opposite to the deviation of the previous measurement from target. For example, when the measurement is 11 (one unit above target), the mean is reduced by one unit before the next measurement is generated. The overcontrol increases the deviations from the target.

Figure 1-12 displays the data without adjustment from Fig. 1-11, except that the measure ments after observation number 50 are increased by two units to simulate the effect of a shift in the mean of the process. When there is a true shift in the mean of a process, an adjustment can be useful. Figure 1-12 also displays the data obtained when one adjustment (a decrease of two units) is applied to the mean after the shift is detected (at observation number 57). Note that this adjustment decreases the deviations from target.

The question of when to apply adjustments (and by what amounts) begins with an under standing of the types of variation that affect a process. The use of a **control charts** is an invaluable way to examine the variability in time-oriented data. Figure 1-13 presents a control chart for the concentration data from Fig. 1-9. The **center line** on the control chart is just the average of the concentration measurements for the first 20 samples (x-= / 91.5 g l) when the process is stable. The **upper control limit** and the **lower control limit** are a pair of statisti cally derived limits that reflect the inherent or natural variability in the process. These limits are located 3 standard deviations of the concentration values above and below the center line. If the process is operating as it should without any external sources of variability present in the system, the concentration measurements should fluctuate randomly around the center line, and almost all of them should fall between the control limits.

In the control chart of Fig. 1-13, the visual frame of reference provided by the center line and the control limits indicates that some upset or disturbance has affected the process around

FIGURE 1-10 Deming's funnel experiment.

Target Marbles

16

14

12

10

4

With adjustment

FIGURE 1-11 Adjustments applied to random disturbances overcontrol the process and increase the deviations from the target.

2

1 11 21 31 41 51 61 71 81 91 Observation number

0

8 y 6

Without adjustment

sample 20 because all of the following observations are below the center line, and two of them actually fall below the lower control limit. This is a very strong signal that corrective action is required in this process. If we can find and eliminate the underlying cause of this upset, we can improve process performance considerably. Thus control limits serve as decision rules about actions that could be taken to improve the process.

Furthermore, Deming pointed out that data from a process are used for different types of conclusions. Sometimes we collect data from a process to evaluate current production. For example, we might sample and measure resistivity on three semiconductor wafers selected from a lot and use this information to evaluate the lot. This is called an **enumerative study**. However, in many cases, we use data from current production to evaluate future production. We apply conclusions to a conceptual, future population. Deming called this an **analytic study**. Clearly this requires an assumption of a **stable** process, and Deming emphasized that control charts were needed to justify this assumption. See Fig. 1-14 as an illustration.

The use of control charts is a very important application of statistics for monitoring, control ling, and improving a process. The branch of statistics that makes use of control charts is called **statistical process control**, or **SPC**. We will discuss SPC and control charts in Chapter 15.

16

Process mean shift is detected.

units) reduces the

14

12

FIGURE 1-12 Process mean shift is detected at observation number 57, and one adjustment (a decrease of two

10

8 y 6

4

2

Without adjustment With adjustment

0

1 11 21 31 41 51 61 71 81 91 Observation number

deviations from target.

100 90

Upper control limit = 100.5

Time

Acetone concentration

x = 91.50

Population

$x_1, x_2, \ldots, x_n$

Sample Sample $x_1, x_2, \ldots, x_n$

0 20

80

Lower control limit = 82.54

1

1 30

?

Observation number (hour)

Future population ?

FIGURE 1-13 A control chart for the chemical

Enumerative study

Analytic study

5 1 10 5 25

FIGURE 1-14 Enumerative versus analytic study.

process concentration data.

1-3 Mechanistic and Empirical Models

Models play an important role in the analysis of nearly all engineering problems. Much of the formal education of engineers involves learning about the models relevant to specific fields and the techniques for applying these models in problem formulation and solution. As a simple example, suppose that we are measuring the flow of current in a thin copper wire. Our model for this phenomenon might be Ohm's law:

$$\text{Current} = \text{Voltage/Resistance}$$

or

$$I = E/R \quad (1\text{-}2)$$

We call this type of model a **mechanistic model** because it is built from our underlying knowledge of the basic physical mechanism that relates these variables. However, if we performed this measurement process more than once, perhaps at different times, or even on different days, the observed current could differ slightly because of small changes or variations in factors that are not completely controlled, such as changes in ambient temperature, fluctuations in performance of the gauge, small impurities present at different locations in the wire, and drifts in the voltage source. Consequently, a more realistic model of the observed current might be

$$I = E/R + \epsilon \quad (1\text{-}3)$$

Mechanistic and Empirical Models

where ϵ is a term added to the model to account for the fact that the observed values of current flow do not perfectly conform to the mechanistic model. We can think of ϵ as a term that includes the effects of all unmodeled sources of variability that affect this system.

Sometimes engineers work with problems for which no simple or well-understood mechanistic model explains the phenomenon. For instance, suppose that we are interested in the number average molecular weight (M_n) of a polymer. Now we know that M_n is related to the viscosity of the material (V), and it also depends on the amount of catalyst (C) and the temperature (T) in the polymerization reactor when the material is manufactured. The relationship between M_n and these variables is

$$M_n = f(V, C, T) \quad (1\text{-}4)$$

say, where the *form* of the function f is unknown. Perhaps a working model could be developed from a first-order Taylor series expansion, which would produce a model of the form

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T \quad (1\text{-}5)$$

where the β's are unknown parameters. Now just as in Ohm's law, this model will not exactly describe the phenomenon, so we should account for the other sources of variability that may affect the molecular weight by adding another term to the model; therefore,

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T + \epsilon \quad (1\text{-}6)$$

is the model that we will use to relate molecular weight to the other three variables. This type of model is called an **empirical model**; that is, it uses our engineering and scientific knowledge of the phenomenon, but it is not directly developed from our theoretical or first principles understanding of the underlying mechanism.

To illustrate these ideas with a specific example, consider the data in Table 1-2, which contains data on three variables that were collected in an observational study in a semiconductor manu facturing plant. In this plant, the finished semiconductor is wire-bonded to a frame. The variables reported are pull strength (a measure of the amount of force required to break the bond), the wire length, and the height of the die. We would like to find a model relating pull strength to wire length and die height. Unfortunately, there is no physical mechanism that we can easily apply here, so it does not seem likely that a mechanistic modeling approach will be successful.

Figure 1-15 presents a three-dimensional plot of all 25 observations on pull strength, wire length, and die height. From examination of this plot, we see that pull strength increases as both wire length and die height increase. Furthermore, it seems reasonable to think that a model such as Pull strength wire length die height =β +β +β + 01 2 ( )( ) e

would be appropriate as an empirical model for this relationship. In general, this type of empiri cal model is called a **regression model**. In Chapters 11 and 12 we show how to build these models and test their adequacy as approximating functions. We will use a method for estimating the parameters in regression models, called the method of **least squares**, that traces its origins to work by Karl Gauss. Essentially, this method chooses the parameters in the empirical model (the β's) to minimize the sum of the squared distances in each data point and the plane represented by the model equation. Applying this technique to the data in Table 1-2 results in

Pull Strength wire length die height = 2 26 2 74 0 0125 .. . + + ( )( ) (1-7)

where the "hat," or circumflex, over pull strength indicates that this is an estimated or predicted quality.

Figure 1-16 is a plot of the predicted values of pull strength versus wire length and die height obtained from Equation 1-7. Notice that the predicted values lie on a plane above the wire length–die height space. From the plot of the data in Fig. 1-15, this model does not appear unreasonable. The empirical model in Equation 1-7 could be used to predict values of pull strength for various combinations of wire length and die height that are of interest. Essentially, an engineer could use the empirical model in exactly the same way as a mechanistic model.

1-4 Probability and Probability Models

Section 1-1 mentioned that decisions often need to be based on measurements from only a subset of objects selected in a sample. This process of reasoning from a sample of objects to

80

dimensional plot of

40

20

Pull strength 60

FIGURE 1-15 Three the wire bond pull

300400500600

12 8 4 0 0

D

Wire length ie height 200

strength data.

16 20

100 0

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50

2	24.45	8	110
3	31.75	11	120
4	35.00	10	550
5	25.02	8	295
6	16.86	4	200
7	14.38	2	375
8	9.60	2	52
9	24.35	9	100
10	27.50	8	300
11	17.08	4	412
12	37.00	11	400
13	41.95	12	500
14	11.66	2	360
15	21.65	4	205
16	17.89	4	400
17	69.00	20	600
18	10.30	1	585
19	34.93	10	540
20	46.59	15	250
21	44.88	15	290
22	54.12	16	510
23	56.63	17	590
24	22.13	6	100
25	21.15	5	400

conclusions for a population of objects was referred to as *statistical inference*. A sample of three wafers selected from a large production lot of wafers in semiconductor manufacturing was an example mentioned. To make good decisions, an analysis of how well a sample represents a pop ulation is clearly necessary. If the lot contains defective wafers, how well will the sample detect these defective items? How can we quantify the criterion to "detect well?" Basically, how can we quantify the risks of decisions based on samples? Furthermore, how should samples be selected to provide good decisions—ones with acceptable risks? **Probability models** help quantify the risks involved in statistical inference, that is, the risks involved in decisions made every day.

80
of predicted
values of
40
Pull strength
20
60
FIGURE 1-16 Plot
300400500600
12 8 4 0 0
pull strength from the
D
Wire length ie height 200
empirical model. 16 20
100 0

More details are useful to describe the role of probability models. Suppose that a produc tion lot contains 25 wafers. If all the wafers are defective or all are good, clearly any sample will generate all defective or all good wafers, respectively. However, suppose that only 1 wafer in the lot is defective. Then a sample might or might not detect (include) the wafer. A prob ability model, along with a method to select the sample, can be used to quantify the risks that the defective wafer is or is not detected. Based on this analysis, the size of the sample might be increased (or decreased). The risk here can be interpreted as follows. Suppose that a series of lots, each with exactly one defective wafer, is sampled. The details of the method used to select the sample are postponed until randomness is discussed in the next chapter. Neverthe less, assume that the same size sample (such as three wafers) is selected in the same manner from each lot. The proportion of the lots in which the defective wafer are included in the sam ple or, more specifically, the limit of this proportion as the number of lots in the series tends to infinity, is interpreted as the probability that the defective wafer is detected.

A probability model is used to calculate this proportion under reasonable assumptions for the manner in which the sample is selected. This is fortunate because we do not want to attempt to sample from an infinite series of lots. Problems of this type are worked in Chapters 2 and 3. More importantly, this probability provides valuable, quantitative information regard ing any decision about lot quality based on the sample.

Recall from Section 1-1 that a population might be conceptual, as in an analytic study that applies statistical inference to future production based on the data from current pro duction. When populations are extended in this manner, the role of statistical inference and the associated probability models become even more important.

In the previous example, each wafer in the sample was classified only as defective or not. Instead, a continuous measurement might be obtained from each wafer. In Section 1-2.5, concentration measurements were taken at periodic intervals from a production process. Figure 1-8 shows that variability is present in the measurements, and there might be concern that the process has moved from the target setting for concentration. Similar to the defective wafer, one might want to quantify our ability to detect a process change based on the sample data. Control limits were mentioned in Section 1-2.5 as decision rules for whether or not to adjust a process. The probability that a particular process change is detected can be calculated with a probability model for concentration measurements. Models for continuous measurements are developed based on plausible assumptions for the data and a result known as the *central limit theorem*, and the associated normal dis tribution is a particularly valuable probability model for statistical inference. Of course, a check of assumptions is important. These types of probability models are discussed in Chapter 4. The objective is still to quantify the risks inherent in the inference made from the sample data.

Throughout Chapters 6 through 15, we base decisions on statistical inference from sample data. We use continuous probability models, specifically the normal distribution, extensively to quantify the risks in these decisions and to evaluate ways to collect the data and how large a sample should be selected.

Analytic study
Cause and effect
Designed
experiment
Empirical model
Engineering method
Enumerative study

Factorial experiment
Fractional factorial
experiment
Hypothesis
Hypothesis
testing Interaction
Mechanistic model
Observational

study
Overcontrol
Population
Probability model
Random variable
Randomization
Retrospective
study Sample

Scientific method
Statistical inference
Statistical process
control Statistical
thinking
Tampering
Time series
Variability

Chapter Outline

2

Probability

An athletic woman in her twenties arrives at the emergency department complaining of dizziness after running in hot weather. An electrocardiogram is used to check for a heart attack, and the patient generates an abnormal result. The test has a false positive rate 0.1 (the probability of an abnormal result when the patient is normal) and a false negative rate of 0.1 (the probability of a normal result when the patient is abnormal). Furthermore, it might be assumed that the prior probability of a heart attack for this patient is 0.001. Although the abnormal test is a concern, you might be surprised to learn that the probability of a heart attack given the electro cardiogram result is still less than 0.01. See "Why Clinicians are Natural Bayesians" (2005, bmj.com) for details of this example and others.
The key is to properly combine the given probabilities.

Furthermore, the exact same analysis used for this medical example can be applied to tests of engineered products. Con sequently, knowledge of how to manipulate probabilities in order to assess risks and make better decisions is important throughout scientifi c and engineering disciplines. In this chapter, the laws of probability are presented and used to assess risks in cases such as this one and numerous others.

Learning Objectives

After careful study of this chapter, you should be able to do the following:

1. Understand and describe sample spaces and events for random experiments with graphs, tables, lists, or tree diagrams
2. Interpret probabilities and use the probabilities of outcomes to calculate probabilities of events in discrete sample spaces
3. Use permuations and combinations to count the number of outcomes in both an event and the sample space
4. Calculate the probabilities of joint events such as unions and intersections from the probabilities of individual events
5. Interpret and calculate conditional probabilities of events
6. Determine the independence of events and use independence to calculate probabilities 7. Use

Bayes' theorem to calculate conditional probabilities

8. Understand random variables

2-1 Sample Spaces and Events

2-1.1 RANDOM EXPERIMENTS

If we measure the current in a thin copper wire, we are conducting an experiment. However, day-to-day repetitions of the measurement can differ slightly because of small variations in variables that are not controlled in our experiment, including changes in ambient temperatures, slight variations in the gauge and small impurities in the chemical composition of the wire (if different locations are selected), and current source drifts. Consequently, this experiment (as well as many we conduct) is said to have a **random** component. In some cases, the ran

dom variations are small enough, relative to our experimental goals, that they can be ignored. However, no matter how carefully our experiment is designed and conducted, the variation is almost always present, and its magnitude can be large enough that the important conclusions from our experiment are not obvious. In these cases, the methods presented in this book for modeling and analyzing experimental results are quite valuable.

Our goal is to understand, quantify, and model the type of variations that we often encounter. When we incorporate the variation into our thinking and analyses, we can make informed judgments from our results that are not invalidated by the variation.

Models and analyses that include variation are not different from models used in other areas of engineering and science. Fig. 2-1 displays the important components. A math ematical model (or abstraction) of the physical system is developed. It need not be a per fect abstraction. For example, Newton's laws are not perfect descriptions of our physical universe. Still, they are useful models that can be studied and analyzed to approximately quantify the performance of a wide range of engineered products. Given a mathematical abstraction that is validated with measurements from our system, we can use the model to understand, describe, and quantify important aspects of the physical system and predict the response of the system to inputs.

Throughout this text, we discuss models that allow for variations in the outputs of a sys tem, even though the variables that we control are not purposely changed during our study. Fig. 2-2 graphically displays a model that incorporates uncontrollable inputs (noise) that combine with the controllable inputs to produce the output of our system. Because of the uncontrollable inputs, the same settings for the controllable inputs do not result in identical outputs every time the system is measured.

Physical system

Measurements Analysis Model

FIGURE 2-1 Continuous iteration between model and physical system.

Random

Controlled
variables

Input System Output

Noise
variables

FIGURE 2-2 Noise variables affect the transformation of inputs to outputs.

Experiment

An experiment that can result in different outcomes, even though it is repeated in the same manner every time, is called a **random experiment**.

For the example of measuring current in a copper wire, our model for the system might simply be Ohm's law. Because of uncontrollable inputs, variations in measurements of cur rent are expected. Ohm's law might be a suitable approximation. However, if the variations are large relative to the intended use of the device under study, we might need to extend our model to include the variation. See Fig. 2-3.

As another example, in the design of a communication system, such as a computer or voice communication network, the information capacity available to serve individuals using the net work is an important design consideration. For voice communication, sufficient external lines need to be available to meet the requirements of a business. Assuming each line can carry only a single conversation, how many lines should be purchased? If too few lines are purchased, calls can be delayed or lost. The purchase of too many lines increases costs. Increasingly, design and product development is required to meet customer requirements *at a competitive cost*.

In the design of the voice communication system, a model is needed for the number of calls and the duration of calls. Even knowing that, on average, calls occur every five minutes and that they last five minutes is not sufficient. If calls arrived precisely at five-minute intervals and lasted for precisely five minutes, one phone line would be sufficient. However, the slight est variation in call number or duration would result in some calls being blocked by others. See Fig. 2-4. A system designed without considering variation will be woefully inadequate for practical use. Our model for the number and duration of calls needs to include variation as an integral component.

2-1.2 SAMPLE SPACES

To model and analyze a random experiment, we must understand the set of possible **outcomes** from the experiment. In this introduction to probability, we use the basic concepts of sets and operations on sets. It is assumed that the reader is familiar with these topics.

Sample Space

The set of all possible outcomes of a random experiment is called the **sample space** of the experiment. The sample space is denoted as *S*.

A sample space is often defined based on the objectives of the analysis. The following example illustrates several alternatives.

Call

Call duration Time

0 5 10 15 20 Minutes

123

Call 3 blocked

Current

Call

Call duration Time

0 5 10 15 20 Minutes

Voltage

1234

FIGURE 2-3 A closer examination of the system identifies deviations from the model.

FIGURE 2-4 Variation causes disruptions in the system.

Camera Flash Consider an experiment that selects a cell phone camera and records the recycle time of a flash (the time taken to ready the camera for another flash). The possible values for this time depend on the resolution of the timer and on the minimum and maximum recycle times. However, because the time is positive it is convenient to define the sample space as simply the positive real line $S = R^+ = \{x \mid x > 0\}$

If it is known that all recycle times are between 1.5 and 5 seconds, the sample space can be

$$S = \{x \mid 1.5 < x < 5\}$$

If the objective of the analysis is to consider only whether the recycle time is low, medium, or high, the sample space can be taken to be the set of three outcomes

$$S = \{low, medium, high\}$$

If the objective is only to evaluate whether or not a particular camera conforms to a minimum recycle time specification, the sample space can be simplified to a set of two outcomes

$$S = \{yes, no\}$$

that indicates whether or not the camera conforms.

It is useful to distinguish between two types of sample spaces.

Discrete and Continuous Sample Spaces

A sample space is **discrete** if it consists of a finite or countable infinite set of outcomes.

A sample space is **continuous** if it contains an interval (either finite or infinite) of real numbers.

In Example 2-1, the choice $S = R^+$ is an example of a continuous sample space, whereas $S = \{yes, no\}$ is a discrete sample space. As mentioned, the best choice of a sample space depends on the objectives of the study. As specifi c questions occur later in the book, appropriate sample spaces are discussed.

times of two cameras are recorded. The exten sion of the positive real line R is to take the sample space to be the positive quadrant of the plane

$$S = \cdot R$$
$$R^{++}$$

Camera Specifi cations Suppose that the recycle

If the objective of the analysis is to consider only whether or not the cameras conform to the manufacturing specifi cations, either camera may or may not conform. We abbreviate *yes* and *no* as *y* and *n*. If the ordered pair *yn* indicates that the fi rst camera conforms and the second does not, the sample space can be represented by the four outcomes:

If we are interested only in the number of conforming cameras in the sample, we might summarize the sample space as S ,, = { } 012

As another example, consider an experiment in which cameras are tested unitl the fl ash recycle time fails to meet the specifi cations. The sample space can be represented as

S *n, yn, yyn, yyyn, yyyyn,* = { } and so forth

and this is an example of a discrete sample space that is countably infi nite.

Sample spaces can also be described graphically with **tree diagrams**. When a sample space can be constructed in several steps or stages, we can represent each of the n_1 ways of completing the fi rst step as a branch of a tree. Each of the ways of completing the second step can be represented as n_2 branches starting from the ends of the original branches, and so forth.

communication system is classifi ed as to whether it is received within the time specifi ed by the system design. If

tree diagram to represent the sample space of possible outcomes. Each message can be received either on time or late. The possible results for three messages can be displayed by eight branches in the tree diagram shown in Fig. 2-5.

Practical Interpretation: A tree diagram can effectively represent a sample space. Even if a tree becomes too large to construct, it can still conceptually clarify the sample space.

Message 1

On time Late

Message 2

On time Late
On time Late

Message 3

On time Late On time Late On time Late

On time Late

FIGURE 2-5 Tree diagram for three messages.

Automobile Options An automobile manufacturer provides vehicles equipped with selected options. Each

With or without an automatic transmission
With or without a sunroof
With one of three choices of a stereo system
With one of four exterior colors

If the sample space consists of the set of all possible vehicle types, what is the number of outcomes in the sample space? The sample space contains 48 outcomes. The tree diagram for the different types of vehicles is displayed in Fig. 2-6.

Transmission

Automatic Manual

Sunroof

Yes No Yes No

Stereo

1 23 1 23 1 23 1 23 Color

FIGURE 2-6 Tree diagram for different types of vehicles with 48 outcomes in the sample space.

Automobile Colors Consider an extension of the automobile manufacturer illustration in the previous example in which another vehicle option is the interior color. There are four choices of interior color: red, black, blue, or brown. However,

With a red exterior, only a black or red interior can be chosen.

With a white exterior, any interior color can be chosen.

With a blue exterior, only a black, red, or blue interior can be chosen.

With a brown exterior, only a brown interior can be chosen.

In Fig. 2-6, there are 12 vehicle types with each exterior color, but the number of interior color choices depends on the exterior color. As shown in Fig. 2-7, the tree diagram can be extended to show that there are 120 different vehicle types in the sample space.

Exterior color Red White Blue Brown

Interior color Black Red

12 3 2 = 24 12 3 4 = 48 12 3 3 = 36 12 3 1 = 12

24 + 48 + 36 + 12 = 120 vehicle types

FIGURE 2-7 Tree diagram for different types of vehicles with interior colors

2-1.3 EVENTS

Often we are interested in a collection of related outcomes from a random experiment. Related outcomes can be described by subsets of the sample space, and set operations can also be applied.

Event

An **event** is a subset of the sample space of a random experiment.

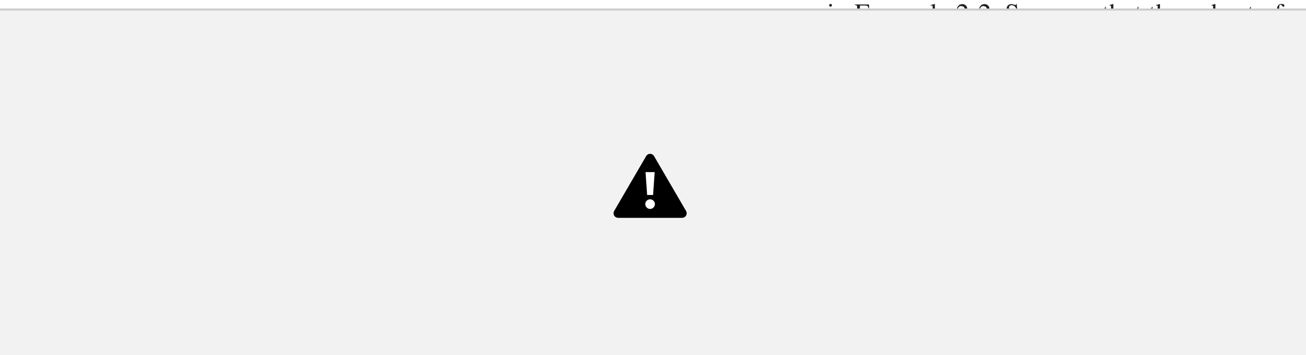
We can also be interested in describing new events from combinations of existing events. Because events are subsets, we can use basic set operations such as unions, intersections, and complements to form other events of interest. Some of the basic set operations are summarized here in terms of events:

outcomes that are contained in either of the two events. We denote the union as $E_1 \cup E_2$.

The **intersection** of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $E_1 \cap E_2$.

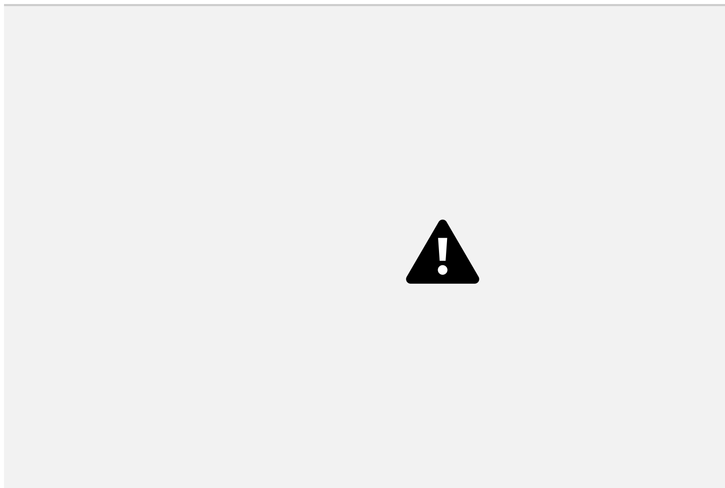
The **complement** of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the complement of the event E as E'. The notation E^C is also used in other literature to denote the complement.

Events Consider the sample space $S = \{yy, yn, ny, nn\}$ i... utcomes for ... d as E_1. Then,



The event such that both cameras do not conform, denoted as E_2, contains only the single outcome, $E_2 = \{nn\}$. Other examples of events are $E_3 = \varnothing$, the null set, and $E_4 = S$, the sample space. If $E_5 = \{yn, ny, nn\}$, $E_1 \cup E_5 = S$, $E_1 \cap E_5 = \{yn, ny\}$, $E_1' = \{nn\}$

Practical Interpretation: Events are used to defi ne outcomes of interest from a random experiment. One is often interested in the probabilities of specifi ed events.



As in Example 2-1, camera recycle times might use the sample space $S = R^+$, the set of posi tive real numbers. Let

$E_1 = \{x \mid 10 \le x < 12\}$ and $E_2 = \{x \mid 11 < x < 15\}$

Then,

$E_1 \cup E_2 = \{x \mid 10 \le x < 15\}$

and

$E_1 \cap E_2 = \{x \mid 11 < x < 12\}$

Also,

$E_1' = \{x \mid x < 10 \text{ or } 12 \le x\}$

and

$E_1' \cap E_2 = \{x \mid 12 \le x < 15\}$

summarizes visits to emergency departments at four hospitals in Arizona. People may leave without being seen denoted as LWBS. The remaining visits are serviced at the emergency department, and the visitor may or may not be admitted for a stay in the hospital.

Let A denote the event that a visit is to hospital 1, and let B denote the event that the result of the visit is LWBS. Calculate the number of outcomes in $A \cap \cup B$, A , A B and $'$.

The event $A \cap B$ consists of the 195 visits to hospital 1 that result in LWBS. The event A' consists of the visits to hospitals 2, 3, and 4 and contains 6991 5640 4329 16 690 $++=$, visits. The event $A \cup B$ consists of the visits to hospital 1 or the visits that result in LWBS, or both, and contains 5292 270 246 6050 $++=$ visits. Notice that the last result can also be calculated as the number of visits in A plus the number of visits in B minus the number of visits $A \cap B$ (that would otherwise be counted twice) = 5292 953 195 6050 $+-=$.

Practical Interpretation: Hospitals track visits that result in LWBS to understand resource needs and to improve patient services.

Hospital

	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

Diagrams are often used to portray relationships between sets, and these diagrams are also used to describe relationships between events. We can use **Venn diagrams** to represent a sample space and events in a sample space. For example, in Fig. 2-8(a) the sample space of the random experiment is represented as the points in the rectangle S. The events A and B are the subsets of points in the indicated regions. Figs. 2-8(b) to 2-8(d) illustrate additional joint events. Fig. 2-9 illustrates two events with no common outcomes.

Mutually Exclusive Events

Two events, denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \varnothing$$

are said to be **mutually exclusive**.

Additional results involving events are summarized in the following. The defi nition of the complement of an event implies that

The distributive law for set operations implies that

$(AB C AC BC AB C AC BC \cup)\cap = \cap ( )\cup \cap ( )$ and

$( ) \cap \cup = \cup ( )\cap \cup ( )$ DeMorgan's laws imply that

$(AB A B AB A B \cup)' = ' \cap ' \cap$ and $( )' = ' \cup '$

Also, remember that

$A \cap = \cap \cup = \cup$ $BBA ABBA$ and

$(E E'' =)$

2-1.4 COUNTING TECHNIQUES

In many of the examples in this chapter, it is easy to determine the number of outcomes in each event. In more complicated examples, determining the outcomes in the sample space (or an event) becomes more diffi cult. Instead, counts of the numbers of outcomes in the sample space and various events are used to analyze the random experiments. These methods are referred to as **counting techniques**. Some simple rules can be used to simplify the calculations.

In Example 2-4, an automobile manufacturer provides vehicles equipped with selected options. Each vehicle is ordered

With or without an automatic transmission
With or without a sunroof
With one of three choices of a stereo system
With one of four exterior colors

$A > B$

Sample space S with events A and B A B

A B (b)

S

S

(a)

A B

$(A < B) > C$

A B

A B
$(A > C)'$

S S

C C

S

(c)

(d)

FIGURE 2-8 Venn diagrams.
FIGURE 2-9 Mutually exclusive events.

Multiplication Rule (for counting techniques)

The tree diagram in Fig. 2-6 describes the sample space of all possible vehicle types. The size of the sample space equals the number of branches in the last level of the tree, and this quantity equals $2 \cdots = 2\ 3\ 4\ 48$. This leads to the following useful result.

Assume an operation can be described as a sequence of k steps, and

the number of ways of completing step 1 is n_1, and
the number of ways of completing step 2 is n_2 for each way of completing step 1, and
the number of ways of completing step 3 is n_3 for each way of completing step 2, and so forth.

The total number of ways of completing the operation is

Web Site Design The design for a Website is to consist of four colors, three fonts, and three positions for an image. From the multiplication rule, $4 \cdot 3 \cdot 3 = 36$ different designs are possible.

Practical

Interpretation: The use of the multipication rule and other counting techniques enables one to easily deter mine the number of outcomes in a sample space or event and this, in turn, allows probabilities of events to be determined.

Permutations

Another useful calculation fi nds the number of ordered sequences of the elements of a set. Consider a set of elements, such as $S = \{a, b, c\}$. A **permutation** of the elements is an ordered sequence of the elements. For example, *abc, acb, bac, bca, cab*, and *cba* are all of the permutations of the elements of *S*.

The number of **permutations** of *n* different elements is *n*! where

$$n! = n(n-1)(n-2)\cdots 2 \cdot 1 \quad \text{(2-1)}$$

selected from a set of *n* different elements is

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} \quad \text{(2-2)}$$

possible?

Permutations of Subsets

This result follows from the multiplication rule. A permutation can be constructed by selecting the element to be placed in the fi rst position of the sequence from the *n* ele ments, then selecting the element for the second position from the remaining $n - 1$ ele ments, then selecting the element for the third position from the remaining $n - 2$ elements, and so forth. Permutations such as these are sometimes referred to as *linear permutations*.

In some situations, we are interested in the number of arrangements of only some of the elements of a set. The following result also follows from the multiplication rule.

Printed Circuit Board A printed circuit board has eight different locations in which a component can be placed. If four different components are to be placed on the board, how many different designs are

The number of permutations of subsets of *r* elements

Each design consists of selecting a location from the eight locations for the fi rst component, a location from the remaining seven for the second component, a location from the remaining six for the third component, and a location from the remaining fi ve for the fourth component. Therefore,

!

P_4^8 8765 ^{8} 4

=⋯=

!

ordered sequences for objects that are not all different. The following result is a useful, general calculation.

The number of permutations of $n = nn\ n_{1\,2} + + \cdots + {}_r$ objects of which n_1 are of one type, n_2 are of a second type, … , and n_r are of an rth type is

Permutations of Similar Objects

different designs are possible

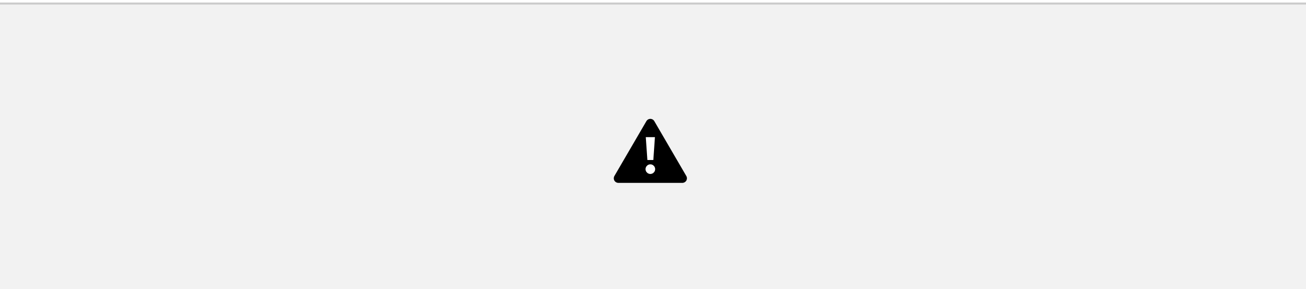
= .

1680

Sometimes we are interested in counting the number of

n !

Hospital Schedule A hospital operating room needs to schedule three knee surgeries and two hip surgeries in a



number of possible sequences of three knee and two hip surgeries is

!

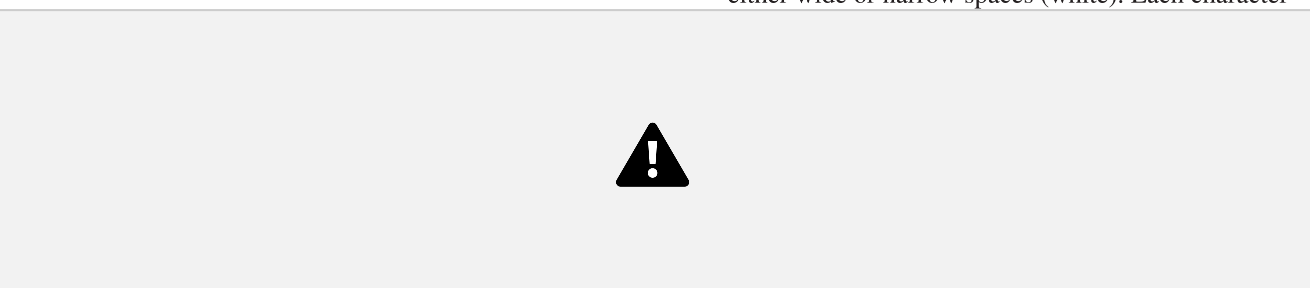
2 310 ! ! =

5

The 10 sequences are easily summarized:

$\{kkkhh,,,,,,,,, kkhkh\ kkhhk\ khkkh\ khhkk\ khhkk\ hkkkh\ hkkhk\ hkhkk\ hhkkk\}$

either wide or narrow spaces (white). Each character



(fi ve

bars and four spaces). The code for a character starts and ends with a bar (either narrow or wide) and a (white) space appears between each bar. The original specifi cation (since revised) used *exactly* two wide bars and one wide space in each character. For example, if b and B denote narrow and wide (black) bars, respectively, and w and W denote narrow and wide (white) spaces, a valid character is $bwBwBWbwb$ (the number 6). One character is held back as a start and stop delimiter. How many other characters can be coded by this system? Can you explain the name of the system?

The four white

spaces occur between the five black bars. In the first step, focus on the bars. The number of permutations of five black bars when two are B and three are b is

$$\frac{5!}{2!\,3!} = 10$$

In the second step, consider the white spaces. A code has three narrow spaces w and one wide space W so there are four possible locations for the wide space. Therefore, the number of possible codes is $10 \cdot 4 = 40$. If one code is held back as a start/stop delimiter, then 39 other characters can be coded by this system (and the name comes from this result).

Combinations

Another counting problem of interest is the number of subsets of r elements that can be selected from a set of n elements. Here, order is not important. These are called **combinations**. Every subset of r elements can be indicated by listing the elements in the set and marking each element with a "$*$" if it is to be included in the subset. Therefore, each permutation of r $*$'s and $n - r$ blanks indicates a different subset, and the numbers of these are obtained from Equation 2-3. For example, if the set is $S = \{a, b, c, d\}$, the subset $\{a, c\}$ can be indicated as

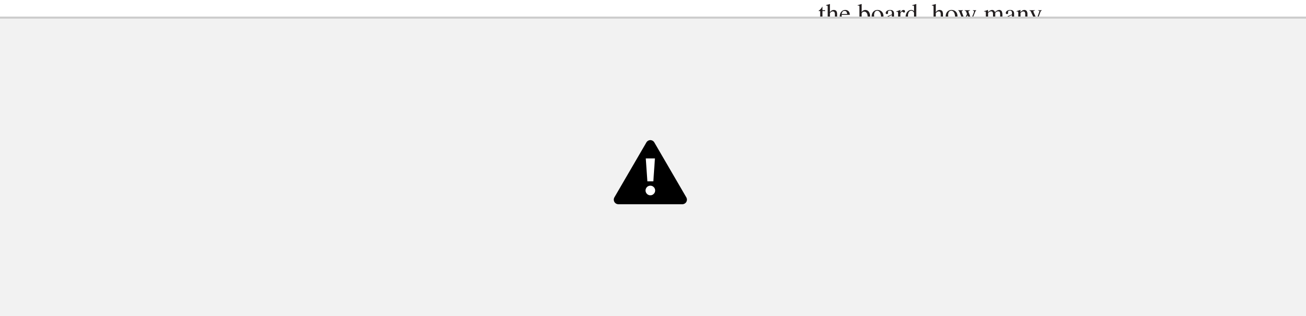
$$\begin{array}{cccc} a & b & c & d \\ * & & * & \end{array}$$

Combinations

The number of combinations, subsets of r elements that can be selected from a set of n elements, is denoted as $\binom{n}{r}$ or C_r^n and

$$C_r^n = \binom{n}{r} = \frac{n!}{r!\,(n-r)!} \qquad (2\text{-}4)$$

has eight different locations in which a component can be placed. If five identical components are to be placed on the board, how many



different designs are possible? Each design is a subset of size five from the eight locations that are to contain the components. From Equation 2-4, the number of possible designs is

$$\frac{8!}{5!\,3!} = 56$$

The following example uses the multiplication rule in combination with Equation 2-4 to answer a more difficult, but common, question. In random experiments in which items are selected from a batch, an item may or may not be replaced before the next one is selected. This is referred to as sampling **with** or **without replacement**, respectively.

manufactured parts contains 3 defective parts and 47 ted from

is, each part can be selected only once, and the sample is a subset of the 50 parts. How many different samples are there of size 6 that contain exactly 2 defective parts?

A subset containing exactly 2 defective parts can be formed by first choosing the 2 defective parts from the three defective parts. Using Equation 2-4, this step can be completed in

$$\binom{3}{2} = \frac{3!}{2!\,1!} = 3 \text{ different ways}$$

Then, the second step is to select the remaining 4 parts from the 47 acceptable parts in the bin. The second step can be completed in

47

Therefore, from the multiplication rule, the number of subsets of size 6 that contain exactly 2 defective parts

is $3 \cdot 178{,}365 = 535{,}095$

As an additional computation, the total number of different subsets of size 6 is found to be

$$\binom{50}{6} = \frac{50!}{6!\,44!} = 15{,}890{,}700$$

SECTION 2-1

FOR

Problem available in *WileyPLUS* at instructor's discretion.

Tutoring problem available in *WileyPLUS* at instructor's discretion

Provide a reasonable description of the sample space for each of the random experiments in Exercises 2-1 to 2-17. There can be more than one acceptable inter pretation of each experiment. Describe any assumptions you make.

2-1. Each of three machined parts is classifi ed as either above or below the target specifi cation for the part.
2-2. Each of four transmitted bits is classifi ed as either in error or not in error.
2-3. In the fi nal inspection of electronic power supplies, either units pass, or three types of nonconformities might occur: func tional, minor, or cosmetic. Three units are inspected. 2-4. The number of hits (views) is recorded at a high-volume Web site in a day.
2-5. Each of 24 Web sites is classifi ed as containing or not containing banner ads.
2-6. An ammeter that displays three digits is used to measure current in milliamperes.
2-7. A scale that displays two decimal places is used to meas ure material feeds in a chemical plant in tons.
2-8. The following two questions appear on an employee survey questionnaire. Each answer is chosen from the fi ve point scale 1 (never), 2, 3, 4, 5 (always).

Is the corporation willing to listen to and fairly evaluate new ideas?

How often are my coworkers important in my overall job performance?

2-9. The concentration of ozone to the nearest part per billion.
2-10. The time until a service transaction is requested of a computer to the nearest millisecond.
2-11. The pH reading of a water sample to the nearest tenth of a unit.
2-12. The voids in a ferrite slab are classifi ed as small, medium, or large. The number of voids in each category is measured by an optical inspection of a sample.
2-13. The time of a chemical reaction is recorded to the near est millisecond.
2-14. An order for an automobile can specify either an automatic or a standard transmission, either with or without air condition ing, and with any one of the four colors red, blue, black, or white.
Describe the set of possible orders for this experiment. 2-15. A sampled injection-molded part could have been pro duced in either one of two presses and in any one of the eight cavities in each press.
2-16. An order for a computer system can specify memory of 4, 8, or 12 gigabytes and disk storage of 200, 300, or 400 gigabytes. Describe the set of possible orders.
2-17. Calls are repeatedly placed to a busy phone line until a connection is achieved.
2-18. Three attempts are made to read data in a magnetic storage device before an error recovery procedure that reposi tions the magnetic head is used. The error recovery procedure attempts three repositionings before an "abort" message is sent to the operator. Let

s denote the success of a read operation
f denote the failure of a read operation
S denote the success of an error recovery procedure
F denote the failure of an error recovery procedure
A denote an abort message sent to the operator

Describe the sample space of this experiment with a tree diagram. 2-19. Three events are shown on the Venn diagram in the fol lowing fi gure:

A B

C

Reproduce the fi gure and shade the region that corresponds to each of the following events.

(a) A' (b) $A \cap B$ (c) $(AB\ C \cap) \cup$ (d) $(B\ C \cup)'$ (e) $(AB\ C \cap)' \cup$

2-20. Three events are shown on the Venn diagram in the fol lowing figure:

A B

C

Reproduce the figure and shade the region that corresponds to each of the following events.

(a) A' (b) $(AB\ AB \cap) \cup \cap (\ ')$ (c) $(AB\ C \cap) \cup$ (d) $(B\ C \cup)'$
(e) $(AB\ C \cap)' \cup$

2-21. A digital scale that provides weights to the nearest gram is used.
(a) What is the sample space for this experiment? Let *A* denote the event that a weight exceeds 11 grams, let *B* denote the event that a weight is less than or equal to 15 grams, and let *C* denote the event that a weight is greater than or equal to 8 grams and less than 12 grams.
Describe the following events.
(b) $A \cup B$ (c) $A \cap B$
(d) A' (e) $A \cup \cup B\ C$
(f) $(A\ C \cup)'$ (g) $A \cap \cap B\ C$
(h) $B\ C' \cap$ (i) $A\ BC \cup \cap (\ )$
2-22. In an injection-molding operation, the length and width, denoted as *X* and *Y*, respectively, of each molded part are eval uated. Let

A denote the event of $48 < X < 52$ centimeters
B denote the event of $9 < Y < 11$ centimeters

Construct a Venn diagram that includes these events. Shade the areas that represent the following:
(a) *A* (b) $A \cap B$
(c) $A' \cup B$ (d) $A\ B \cap$
(e) If these events were mutually exclusive, how successful would this production operation be? Would the process pro duce parts with $X = 50$ centimeters and $Y = 10$ centimeters?
2-23. Four bits are transmitted over a digi tal communications channel. Each bit is either distorted or received without distortion. Let *Ai* denote the event that the *i*th bit is distorted, $i ,, = \ldots 1\ 4$.

(a) Describe the sample space for this experiment. (b) Are the A_i's mutually exclusive? Describe the outcomes in each of the following events: (c) A_1 (d) A_1'

(e) $AAAA_{1234}\cap\cap\cap$ (f) $(AA\,AA_{12\,34}\cap)\cup\cap(\,)$ 2-24. In light-dependent photosynthesis, light quality refers to the wavelengths of light that are important. The wavelength of a sample of photosynthetically active radiations (PAR) is measured to the nearest nanometer. The red range is 675–700 nm and the blue range is 450–500 nm. Let A denote the event

that PAR occurs in the red range, and let B denote the event that PAR occurs in the blue range. Describe the sample space and indicate each of the following events:

(a) A (b) B (c) $A\ B\cap$ (d) $A\cup B$

2-25. In control replication, cells are replicated over a period of two days. Not until mitosis is completed can freshly synthesized DNA be replicated again. Two control mechanisms have been identified—one positive and one negative. Suppose that a replication is observed in three cells. Let A denote the event that all cells are identified as positive, and let B denote the event that all cells are nega tive. Describe the sample space graphically and display each of the following events:

(a) A (b) B

(c) $A\cap B$ (d) $A\cup B$

2-26. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized here:

		Shock Resistance	
		High	Low
Scratch Resistance	High	70	9
	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Determine the number of disks in $A\cap' B$, A, and $A\cup B$. 2-27. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

		Edge Finish	
		Excellent	Good
Surface Finish	Excellent	80	2
	Good	10	8

(a) Let A denote the event that a sample has excellent sur face finish, and let B denote the event that a sample has excellent edge finish. Determine the number of samples in $A''\cap B$, B and in $A\ B\cup$.

(b) Assume that each of two samples is to be classified on the basis of surface finish, either excellent or good, and on the basis of edge finish, either excellent or good. Use a tree dia gram to represent the possible outcomes of this experiment.

2-28. Samples of emissions from three suppliers are clas sified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
Supplier	1	22	8
	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. Determine the number of samples in $A''\cap B$, B, and $A\cup B$.

2-29. The rise time of a reactor is measured in minutes (and fractions of minutes). Let the sample space be posi tive, real numbers. Define the events A and B as follows: $A\ xx=<.\{\}$ $|\ 72\ 5$ and $B\ xx=>.\{\}|\ 52\ 5$.

Describe each of the following events:

(a) A' (b) B'

(c) $A\cap B$ (d) $A\cup B$

2-30. A sample of two items is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:

(a) The batch contains the items $\{abcd\ ,,,\}$.

(b) The batch contains the items $\{abcde\,f\,g\ ,,,\ ,,\ ,\}$. (c) The batch contains 4 defective items and 20 good items. (d) The batch contains 1 defective item and 20 good items. 2-31. A sample of two printed circuit boards is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:

(a) The batch contains 90 boards that are not defective, 8 boards with minor defects, and 2 boards with major defects.

(b) The batch contains 90 boards that are not defective, 8 boards with minor defects, and 1 board with major defects.

2-32. Counts of the Web pages provided by each of two com puter servers in a selected hour of the day are recorded. Let A denote the event that at least 10 pages are provided by server 1, and let B denote the event that at least 20 pages are provided by server 2. Describe the sample space for the numbers of pages for the two servers graphically in an $x\ y-$ plot. Show each of the following events on the sample space graph:

(a) A (b) B

(c) $A\cap B$ (d) $A\cup B$

2-33. A reactor's rise time is measured in minutes (and frac tions of minutes). Let the sample space for the rise time of each batch be positive, real numbers. Consider the rise times of *two* batches. Let A denote the event that the rise time of batch 1 is less than 72.5 minutes, and let B denote the event that the rise time of batch 2 is greater than 52.5 minutes.

Describe the sample space for the rise time of two batches graphically and show each of the following events on a two dimensional plot:

(a) A (b) B'

(c) $A\cap B$ (d) $A\cup B$

2-34. A wireless garage door opener has a code deter mined by the up or down setting of 12 switches. How many outcomes are in the sample space of possible codes? 2-35. An order for a computer can specify any one of five memory sizes, any one of three types of displays, and any one of four sizes of a hard disk, and can either include or not include a pen tablet. How many different systems can be ordered? 2-36. In a manufacturing operation, a part is produced by machining, polishing, and painting. If there are three machine tools, four polishing tools, and three painting tools, how many different routings (consisting of machining, followed by pol ishing, and followed by painting) for a part are possible? 2-37. New designs for a wastewater treatment tank have proposed three possible shapes, four possible sizes, three loca tions for input valves, and four locations for output valves. How many different product designs are possible?

2-38. A manufacturing process consists of 10 operations that can be completed in any order. How many different pro duction sequences are possible?

2-39. A manufacturing operation consists of 10 opera tions. However, five machining operations must be completed before any of the remaining five assembly operations can begin. Within each set of five, operations can be completed in any order. How many different production sequences are possible? 2-40. In a sheet metal operation, three notches and four bends are required. If the operations can be done in any order, how many different ways of completing the manufacturing are possible?

2-41. A batch of 140 semiconductor chips is inspected by choosing a sample of 5 chips. Assume 10 of the chips do not conform to customer requirements.

(a) How many different samples are possible?

(b) How many samples of five contain exactly one noncon forming chip?

(c) How many samples of five contain at least one noncon forming chip?

2-42. In the layout of a printed circuit board for an electronic product, 12 different locations can accommodate chips. (a) If five different types of chips are to be placed on the board, how many different layouts are possible?

(b) If the five chips that are placed on the board are of the same type, how many different layouts are possible?

2-43. In the laboratory analysis of samples from a chemical process, five samples from the process are analyzed daily. In addition, a control sample is analyzed twice each day to check the calibration of the laboratory instruments.

(a) How many different sequences of process and control sam ples are possible each day? Assume that the five process samples are considered identical and that the two control samples are considered identical.

(b) How many different sequences of process and control sam ples are possible if we consider the five process samples to be different and the two control samples to be identical?

(c) For the same situation as part (b), how many sequences are possible if the first test of each day must be a control sample?

2-44. In the design of an electromechanical product, 12 com ponents are to be stacked into a cylindrical casing in a manner that minimizes the impact of shocks. One end of the casing is designated as the bottom and the other end is the top. (a) If all components are different, how many different designs are possible?

(b) If seven components are identical to one another, but the others are different, how many different designs are possible?

(c) If three components are of one type and identical to one another, and four components are of another type and identical to one another, but the others are different, how many different designs are possible?

2-45. Consider the design of a communication system. (a) How many three-digit phone prefixes that are used to repre sent a particular geographic area (such as an area code) can be created from the digits 0 through 9?

(b) As in part (a), how many three-digit phone prefixes are possible that do not start with 0 or 1, but contain 0 or 1 as the middle digit? (c) How many three-digit phone prefixes are possible in which no digit appears more than once in each prefix?

2-46. A *byte* is a sequence of eight bits and each bit is either 0 or 1. (a) How many different bytes are possible?

(b) If the first bit of a byte is a parity check, that is, the first byte is determined from the other seven bits, how many different bytes are possible?

2-47. In a chemical plant, 24 holding tanks are used for final product storage. Four tanks are selected at random and without replacement. Suppose that six of the tanks contain material in which the viscosity exceeds the customer requirements.

(a) What is the probability that exactly one tank in the sample contains high-viscosity material?

(b) What is the probability that at least one tank in the sample contains high-viscosity material?

(c) In addition to the six tanks with high viscosity levels, four different tanks contain material with high impurities. What is the probability that exactly one tank in the sample con tains high-viscosity material and exactly one tank in the sample contains material with high impurities?

2-48. Plastic parts produced by an injection-molding operation are checked for conformance to specifications. Each tool contains 12 cavities in which parts are produced, and these parts fall into a conveyor when the press opens. An inspector chooses 3 parts from among the 12 at random. Two cavities are affected by a temperature malfunction that results in parts that do not conform to specifications.

(a) How many samples contain exactly 1 nonconforming part? (b) How many samples contain at least 1 nonconforming part? 2-49. A bin of 50 parts contains 5 that are defective. A sample of 10 parts is selected at random, without replacement. How many samples contain at least four defective parts? 2-50. The following table summarizes 204 endothermic reac tions involving sodium bicarbonate.

Final Temperature Conditions	**Heat Absorbed (cal)**

	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is below target. Determine the number of reactions in each of the following events.

(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cup B'$ (e) $A' \cap B'$

2-51. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. How many different designs are possible?

2-52. Consider the hospital emergency department data in Example 2-8. Let A denote the event that a visit is to hospital 1, and let B denote the event that a visit results in admittance to any hospital. Determine the number of persons in each of the following events.

(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cup B'$ (e) $A' \cap B'$

2-53. An article in *The Journal of Data Science* ["A Statistical Analysis of Well Failures in Baltimore County" (2009, Vol. 7, pp. 111–127)] provided the following table of well failures for different geological formation groups in Baltimore County.

Geological Formation Group	Wells Failed	Total
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

Let A denote the event that the geological formation has more than 1000 wells, and let B denote the event that a well failed. Determine the number of wells in each of the following events.

(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cap B'$ (e) $A' \cap B'$

2-54. Similar to the hospital schedule in Example 2-11, suppose that an operating room needs to handle three knee, four hip, and five shoulder surgeries.

(a) How many different sequences are possible?

(b) How many different sequences have all hip, knee, and shoulder surgeries scheduled consecutively?

(c) How many different schedules begin and end with a knee surgery?

2-55. Consider the bar code in Example 2-12. One code is still held back as a delimiter. For each of the following cases, how many characters can be encoded?

(a) The constraint of exactly two wide bars is replaced with one that requires exactly one wide bar.

(b) The constraint of exactly two wide bars is replaced with one that allows either one or two wide bars.

(c) The constraint of exactly two wide bars is dropped.

(d) The constraints of exactly two wide bars and one wide space are dropped.

2-56. A computer system uses passwords that contain exactly eight characters, and each character is 1 of the 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Let Ω denote the set of all possible passwords, and let A and B denote the events that consist of passwords with only letters or only integers, respectively. Determine the number of passwords in each of the following events.

(a) Ω (b) A (c) $A' \cap B'$

(d) Passwords that contain at least 1 integer

(e) Passwords that contain exactly 1 integer

2-57. The article "Term Efficacy of Ribavirin Plus Interferon Alfa in the Treatment of Chronic Hepatitis C," [*Gastroenterology* (1996, Vol. 111, no. 5, pp. 1307–1312)], considered the effect of two treatments and a control for treatment of hepatitis C. The following table provides the total patients in each group and the number that showed a complete (positive) response after 24 weeks of treatment.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let A denote the event that the patient was treated with ribavirin plus interferon alfa, and let B denote the event that the response was complete. Determine the number of patients in each of the following events.

(a) A (b) $A \cap B$ (c) $A \cup B$ (d) $A' \cap B'$

2-2 Interpretations and Axioms of Probability

In this chapter, we introduce probability for **discrete sample spaces**—those with only a finite (or countably infinite) set of outcomes. The restriction to these sample spaces enables us to simplify the concepts and the presentation without excessive mathematics.

Probability is used to quantify the likelihood, or chance, that an outcome of a random experiment will occur. "The chance of rain today is 30%" is a statement that quantifies our feeling about the possibility of rain. The likelihood of an outcome is quantified by assigning a number from the interval [0, 1] to the outcome (or a percentage from 0 to 100%). Higher

num
bers indicate that the outcome is more likely than lower numbers. A 0 indicates an outcome will not occur. A probability of 1 indicates that an outcome will occur with certainty.

The probability of an outcome can be interpreted as our subjective probability, or **degree of belief**, that the outcome will occur. Different individuals will no doubt assign different proba bilities to the same outcomes. Another interpretation of probability is based on the conceptual model of repeated replications of the random experiment. The probability of an outcome is interpreted as the limiting value of the proportion of times the outcome occurs in *n* repetitions of the random experiment as *n* increases beyond all bounds. For example, if we assign prob ability 0.2 to the outcome that there is a corrupted pulse in a digital signal, we might interpret this assignment as implying that, if we analyze many pulses, approximately 20% of them will be corrupted. This example provides a **relative frequency** interpretation of probability. The proportion, or relative frequency, of replications of the experiment that result in the outcome is 0.2. Probabilities are chosen so that the sum of the probabilities of all outcomes in an experi ment adds up to 1. This convention facilitates the relative frequency interpretation of prob ability. Fig. 2-10 illustrates the concept of relative frequency.

Probabilities for a random experiment are often assigned on the basis of a reasonable model of the system under study. One approach is to base probability assignments on the simple con cept of equally likely outcomes. For example, suppose that we select 1 laser diode **randomly** from a batch of 100. *Randomly* implies that it is reasonable to assume that each diode in the batch has an equal chance of being selected. Because the sum of the probabilities must equal 1, the probability model for this experiment assigns probability of 0.01 to each of the 100 out comes. We can interpret the probability by imagining many replications of the experiment. Each time we start with all 100 diodes and select 1 at random. The probability 0.01 assigned to a particular diode represents the proportion of replicates in which a particular diode is selected. When the model of **equally likely outcomes** is assumed, the probabilities are chosen to be equal.

Corrupted pulse

Voltage

Time

Relative frequency of corrupted pulse = $\frac{2}{10}$

FIGURE 2-10 Relative frequency of corrupted pulses sent over a communication channel.

Equally Likely Outcomes

Whenever a sample space consists of *N* possible outcomes that are equally likely, the probability of each outcome is $1/N$.

events that are composed of several outcomes from the sample space. This is straightforward for a discrete sample space.

Laser Diodes Assume that 30% of the laser diodes in a batch of 100 meet the minimum power requirements of a specifi c customer. If a laser diode is selected randomly, that is, each laser diode is equally likely to be selected, our intuitive feeling is that the probability of meeting the customer's requirements is 0.30. Let E denote the subset of 30 diodes that meet the customer's requirements. Because E contains 30 outcomes and each outcome has probability 0.01, we conclude that the probability of E is 0.3. The conclusion matches our intuition. Fig. 2-11 illustrates this example.

E

Diodes

S

$P(E) = 30(0.01) = 0.30$

FIGURE 2-11 Probability of the event E is the sum of the probabilities of the outcomes in E.

For a discrete sample space, the probability of an event can be defi ned by the reasoning used in the preceding example.

Probability of an Event

For a discrete sample space, the *probability of an event E,* denoted as $P(E)$, equals the sum of the probabilities of the outcomes in E.

Probabilities of Events A random experiment can result in one of the outcomes $\{a, b, c, d\}$ with probabilities 0.1, 0.3, 0.5, and 0.1, respectively. Let A denote the event $\{a, b\}$, B the event $\{b, c, d\}$, and C the event $\{d\}$. Then, $P A$ 03 05 01 09 = . 0 1

( )

$P B$

( )

$P C$

( )

=.+.=.

01 03 04

=.+.+.=.

Also, PA ,PB ( )′ ′= = . 0 6 0 1, and ( ) . $P C$( )′ = . 0 9. Furthermore, because $A∩ = B b$, { } $PA B$ ( ) ∩ = . 0 3. Because $A∪ = B$ $abcd$ {, ,, },$PA B$ ( ). . . . . ∪= + + + = 01 03 05 01 1 Because $A∩C$ is the null set, $PA C$ ( ) ∪ = 0.

Contamination Particles A visual inspection of a location on wafers from a semiconductor manufacturing

If one wafer is selected randomly from this process and the location is inspected, what is the probability that it contains no particles? If information were available for each wafer, we could define the sample space as the set of all wafers inspected and proceed as in the example with diodes. However, this level of detail is not needed in this case. We can consider the sample space to consist of the six categories that summarize the number of contamination particles on a wafer. Each category has probability equal to the proportion of wafers in the category. The event that there is no

0	0.40
1	0.20
2	0.15
3	0.10
4	0.05
5 or more	0.10

contamination particle in the inspected location on the wafer, denoted as *E*, can be considered to be composed of the single outcome, namely, $E = \{0\}$. Therefore,

$$P(E) = 0.4$$

What is the probability that a wafer contains three or more particles in the inspected location? Let *E* denote the event that a wafer contains three or more particles in the inspected location. Then, *E* consists of the three outcomes {3, 4, 5 or more}. Therefore,

$$P(E) = 0.10 + 0.05 + 0.10 = 0.25$$

Practical Interpretation: Contamination levels affect the yield of functional devices in semiconductor manufacturing so that probabilities such as these are regularly studied.

Often more than one item is selected from a batch without replacement when production is inspected. In this case, *randomly* selected implies that each possible subset of items is equally likely.

Manufacturing Inspection Consider the inspection described in Example 2-14. From a bin of 50 parts, 6 parts are selected randomly without replacement. The bin contains 3 defective parts and 47 nondefective parts. What is the probability that exactly 2 defective parts are selected in the sample? The sample space consists of all possible (unordered) subsets of 6 parts selected without replacement. As shown in Example 2-14, the number of subsets of size 6 that contain exactly 2 defective parts is 535,095 and the total number of subsets of size 6 is 15,890,700. The probability of an event is determined as the ratio of the number of outcomes in the event to the number of outcomes in the sample space (for equally likely outcomes). Therefore, the probability that a sample contains exactly 2 defective parts is

$$\frac{535,095}{15,890,700} = 0.034$$

A subset with no defective parts occurs when all 6 parts are selected from the 47 nondefective ones. Therefore, the number of subsets with no defective parts is

$$\binom{47}{6} = \frac{47!}{6!\,41!} = 10,737,573$$

and the probability that no defective parts are selected is

$$\frac{10,737,573}{15,890,700} = 0.676$$

Therefore, the sample of size 6 is likely to omit the defective parts. This example illustrates the hypergeometric distribution studied in Chapter 3.

Now that the probability of an event has been defined, we can collect the assumptions that we have made concerning probabilities into a set of **axioms** that the probabilities in any random experiment must satisfy. The axioms ensure that the probabilities assigned in an experiment can be interpreted as relative frequencies and that the assignments are consistent with our intuitive understanding of relationships between relative frequencies. For example, if event A is contained in event B, we should have $P(A) \le P(B)$. The **axioms do not determine probabilities**; the probabilities are assigned based on our knowledge of the system under study. However, the axioms enable us to easily calculate the probabilities of some events from knowledge of the probabilities of other events.

Axioms of Probability

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties: and for any event E,

If S is the sample space and E is any event in a random experiment,

(1) $P(S) = 1$

(2) $0 \le P(E) \le 1$

(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \varnothing$

$P(E_1 \cup E_2) = P(E_1) + P(E_2)$

The property that $0 \le P(E) \le 1$ is equivalent to the requirement that a relative frequency must be between 0 and 1. The property that $P(S) = 1$ is a consequence of the fact that an outcome from the sample space occurs on every trial of an experiment. Consequently, the relative frequency of S is 1. Property 3 implies that if the events E_1 and E_2 have no outcomes in common, the relative frequency of outcomes in $E_1 \cup E_2$ is the sum of the relative frequencies of the outcomes in E_1 and E_2.

These axioms imply the following results. The derivations are left as exercises at the end of this section. Now,

$P(E') = 1 - P(E)$

For example, if the probability of the event E is 0.4, our interpretation of relative frequency implies that the probability of E' is 0.6. Furthermore, if the event E_1 is contained in the event E_2, $P(E_1) \le P(E_2)$

$P(\varnothing) = 0$

FOR SECTION 2-2

Problem available in *WileyPLUS* at instructor's discretion.

Tutoring problem available in *WileyPLUS* at instructor's discretion

2-58. Each of the possible five outcomes of a random experiment is equally likely. The sample space is $\{a, b, c, d, e\}$. Let A denote the event $\{a, b\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

(a) $P(A)$ (b) $P(B)$ (c) $P(A')$ (d) $P(A \cup B)$ (e) $P(A \cap B)$

2-59. The sample space of a random experiment is $\{a, b, c, d, e\}$ with probabilities 0.1, 0.1, 0.2, 0.4, and 0.2, respectively. Let A denote the event $\{a, b, c\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

(a) $P(A)$ (b) $P(B)$ (c) $P(A')$ (d) $P(A \cup B)$ (e) $P(A \cap B)$

2-60. Orders for a computer are summarized by the optional features that are requested as follows:

Proportion of Orders
No optional features 0.3 One optional feature 0.5 More than one optional feature 0.2

(a) What is the probability that an order requests at least one optional feature?

(b) What is the probability that an order does not request more than one optional feature?

2-61. If the last digit of a weight measurement is equally likely to be any of the digits 0 through 9,

(a) What is the probability that the last digit is 0? (b) What is the probability that the last digit is greater than or equal to 5?

2-62. A part selected for testing is equally likely to have been produced on any one of six cutting tools.

(a) What is the sample space?

(b) What is the probability that the part is from tool 1? (c) What is the probability that the part is from tool 3 or tool 5?

(d) What is the probability that the part is not from tool 4?

2-63. An injection-molded part is equally likely to be obtained from any one of the eight cavities on a mold.

(a) What is the sample space?

(b) What is the probability that a part is from cavity 1 or 2? (c) What is the probability that a part is from neither cavity 3 nor 4? 2-64. In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Because acids and bases are usually colorless (as are the water and salt produced in the neutralization reaction), pH is

measured to monitor the reaction. Suppose that the equiva lence point is reached after approximately 100 mL of an NaOH solution has been added (enough to react with all the acetic acid present) but that replicates are equally likely to indicate from 95 to 104 mL to the nearest mL. Assume that volumes are measured to the nearest mL and describe the sample space. (a) What is the probability that equivalence is indicated at 100 mL?

(b) What is the probability that equivalence is indicated at less than 100 mL?

(c) What is the probability that equivalence is indicated between 98 and 102 mL (inclusive)?

2-65. In a NiCd battery, a fully charged cell is composed of nickelic hydroxide. Nickel is an element that has multiple oxida tion states and that is usually found in the following states:

Nickel Charge Proportions Found
0 0.17
+2 0.35
+3 0.33
+4 0.15

(a) What is the probability that a cell has at least one of the positive nickel-charged options?

(b) What is the probability that a cell is *not* composed of a positive nickel charge greater than +3?

2-66. A credit card contains 16 digits between 0 and 9. How ever, only 100 million numbers are valid. If a number is entered randomly, what is the probability that it is a valid number? 2-67. Suppose your vehicle is licensed in a state that issues license plates that consist of three digits (between 0 and 9) followed by three letters (between *A* and *Z*). If a license number is selected randomly, what is the probability that yours is the one selected?

2-68. A message can follow different paths through servers on a network. The sender's message can go to one of five servers for the first step; each of them can send to five servers at the second step; each of those can send to four servers at the third step; and then the message goes to the recipient's server.

(a) How many paths are possible?

(b) If all paths are equally likely, what is the probability that a message passes through the first of four servers at the third step?

2-69. Magnesium alkyls are used as homogenous catalysts in the production of linear low-density polyethylene (LLDPE), which requires a finer magnesium powder to sustain a reaction. Redox reaction experiments using four different amounts of magnesium powder are performed. Each result may or may not be further reduced in a second step using three different magnesium powder amounts. Each of these results may or may not be further reduced in a third step using three different amounts of magnesium powder.

(a) How many experiments are possible?

(b) If all outcomes are equally likely, what is the probability that the best result is obtained from an experiment that uses all three steps?

(c) Does the result in part (b) change if five or six or seven dif ferent amounts are used in the first step? Explain.

2-70. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

Shock Resistance
High Low
Scratch High 70 9 Resistance Low 16 5

Let *A* denote the event that a disk has high shock resistance, and let *B* denote the event that a disk has high scratch resist ance. If a disk is selected at random, determine the following probabilities:

(a) $P\,A(\ )$ (b) $P\,B(\ )$ (c) $P\,A(\ ')$ (d) $PA\ B\ (\cap\ )$ (e) $PA\ B\ (\cup\ )$ (f) $PA\ B\ ('\cup\ )$ 2-71. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

Conforms

	Yes No
	1 22 8
Supplier 2 25 5 3 30 10	

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. If a sam ple is selected at random, determine the following probabilities: (a) $P(A)$ (b) $P(B)$ (c) $P(A)'$ (d) $P(A \cap B)$ (e) $P(A \cup B)$ (f) $P(A' \cap B)$

2-72. An article in the *Journal of Database Management* [“Experimental Study of a Self-Tuning Algorithm for DBMS Buffer Pools” (2005, Vol. 16, pp. 1–20)] provided the workload used in the TPC-C OLTP (Transaction Processing Performance Council’s Version C On-Line Transaction Processing) bench mark, which simulates a typical order entry application. See Table 2E-1.

The frequency of each type of transaction (in the second column) can be used as the percentage of each type of trans action. The average number of *selects* operations required for each type of transaction is shown. Let A denote the event of transactions with an average number of *selects* operations of 12 or fewer. Let B denote the event of transactions with an aver age number of *updates* operations of 12 or fewer. Calculate the following probabilities.

(a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$ (d) $P(A \cap B')$
(e) $P(A \cup B)$

2-73. Use the axioms of probability to show the following: (a) For any event E, $P(E') = 1 - P(E)$.
(b) $P(\varnothing) = 0$
(c) If A is contained in B, then $P(A) \leq P(B)$.

2-74. Consider the endothermic reaction’s in Exercise 2-50. Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is above target. Determine the following probabilities.

2E-1 Average Frequencies and Operations in TPC-C

	Joins
New order	43
23	11
12000	
Payment	
44 4.2	3

1 0 0.6 0 Order status 4 11.4 0 0 0 0.6 0 Delivery 5 130 120 0 10 0 0 Stock level 4 0 0 0 0 0 1

(a) $P(A \cap B)$ (b) $P(A')$ (c) $P(A \cup B)$ (d) $P(A \cup B')$ (e) $P(A' \cap B')$

2-75. A Web ad can be designed from four different colors, three font types, fi ve font sizes, three images, and fi ve text phrases. A specifi c design is randomly generated by the Web server when you visit the site. If you visit the site fi ve times, what is the probability that you will not see the same design?

2-76. Consider the hospital emergency room data in Example 2-8. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Determine the following probabilities.
(a) $P(A \cap B)$ (b) $P(A)'$ (c) $P(A \cup B)$ (d) $P(A \cup B')$ (e) $P(A' \cap B')$

2-77. Consider the well failure data in Exercise 2-53. Let A denote the event that the geological formation has more than 1000 wells, and let B denote the event that a well failed. Deter mine the following probabilities.
(a) $P(A \cap B)$ (b) $P(A)'$ (c) $P(A \cup B)$ (d) $P(A \cup B')$ (e) $P(A' \cap B')$

2-78. Consider the bar code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Determine the probability for each of the following: (a) A wide space occurs before a narrow space. (b) Two wide bars occur consecutively.
(c) Two consecutive wide bars are at the start or end.
(d) The middle bar is wide.

2-3 Addition Rules

2-79. Similar to the hospital schedule in Example 2-11, suppose that an oper ating room needs to schedule three knee, four hip, and fi ve shoulder surgeries. Assume that all schedules are equally likely. Determine the probability for each of the following:
(a) All hip surgeries are completed before another type of surgery. (b) The schedule begins with a hip surgery.
(c) The fi rst and last surgeries are hip surgeries. (d) The fi rst two surgeries are hip surgeries.

2-80. Suppose that a patient is selected randomly from the those described in Exercise 2-57. Let A denote the event that the patient is in the group treated with interferon alfa, and let B denote the event that the patient has a complete response. Determine the following probabilities.

(a) $P(A)$ (b) $P(B)$

(c) $P(A \cap B)$ (d) $P(A \cup B)$ (e) $P(A' \cup B)$

2-81. A computer system uses passwords that contain exactly eight characters, and each character is one of 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Let Ω denote the set of all possible passwords, and let A and B denote the events that consist of passwords with only letters or only integers, respectively. Suppose that all passwords in Ω are equally likely. Determine the probability of each of the following:

(a) A (b) B
(c) A password contains at least 1 integer.
(d) A password contains exactly 2 integers.

Joint events are generated by applying basic set operations to individual events. Unions of events, such as $A \cup B$; intersections of events, such as $A \cap B$; and complements of events, such as A'—are commonly of interest. The probability of a joint event can often be determined from the probabilities of the individual events that it comprises. Basic set operations are also sometimes helpful in determining the probability of a joint event. In this section, the focus is on unions of events.

Semiconductor Wafers Table 2-1 lists the history of 940 wafers in a semiconductor manufacturing process. Suppose that 1 wafer is selected at random. Let H denote the event that the wafer contains high levels of contamination. Then, $P(H) = 358/940$.

Let C denote the event that the wafer is in the center of a sputtering tool. Then, $P(C) = 626/940$. Also, $P(H \cap C)$ is the probability that the wafer is from the center of the sputtering tool and contains high levels of contamination. Therefore,

The event $H \cup C$ is the event that a wafer is from the center of the sputtering tool or contains high levels of contamination (or both). From the table, $P(H \cup C) = 872/940$. An alternative calculation of $P(H \cup C)$ can be obtained as follows. The 112 wafers in the event $H \cap C$ are included once in the calculation of $P(H)$ and again in the calculation of $P(C)$. Therefore, $P(H \cup C)$ can be determined to be

$$P(H \cup C) = P(H) + P(C) - P(H \cap C)$$

$$= 358/940 + 626/940 - 112/940 = 872/940$$

Practical Interpretation: To better understand the sources of contamination, yield from defferent locations on wafers are routinely aggregated.

2-1 Wafers in Semiconductor Manufacturing Classifi ed by Contamination and Location

Location in Sputtering Tool			
Contamination	**Center**	**Edge**	**Total**
Low	514	68	582
High	112	246	358
Total	626	314	

The preceding example illustrates that the probability of A or B is interpreted as $P(A \cup B)$ and that the following general **addition rule** applies.

Probability of a Union

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (2\text{-}5)$$

wafers in Example 2-19 were further classifi ed by the degree of contamination. Table 2-2 shows the proportion of wafers in each category. What is the probability that a wafer was either at the edge or that it contains four or more particles? Let E_1 denote the event that a wafer contains four or more particles, and let E_2 denote the event that a wafer was at the edge.

The requested probability is $P(E_1 \cup E_2)$. Now, $P(E_1) = 0.15$ and $P(E_2) = 0.28$. Also, from the table, $P(E_1 \cap E_2) = 0.04$. Therefore, using Equation 2-1, we fi nd that

$$P(E_1 \cup E_2) = 0.15 + 0.28 - 0.04 = 0.39$$

2-2 Wafers Classifi ed by Contamination and Location

Number of Contamination Particles	Center	Edge	Totals
0	0.30	0.10	0.40
1	0.15	0.05	0.20
2	0.10	0.05	0.15
3	0.06	0.04	0.10
4	0.04	0.01	0.05
5 or more	0.07	0.03	0.10
Totals	0.72	0.28	1.00

What is the probability that a wafer contains less than two particles or that it is both at the edge and contains more than four particles? Let E_1 denote the event that a wafer contains less than two particles, and let E_2 denote the event that a wafer is both at the edge and contains more than four particles. The requested probability is $P(E_1 \cup E_2)$. Now, $P(E_1) = 0.60$ and $P(E_2) = 0.03$. Also, E_1 and E_2 are mutually exclusive. Consequently, there are no wafers in the intersection and $P(E_1 \cap E_2) = 0$. Therefore,

$$P(E_1 \cup E_2) = 0.60 + 0.03 = 0.63$$

Recall that two events A and B are said to be mutually exclusive if $A \cap B = \varnothing$. Then, $P(A \cap B) = 0$, and the general result for the probability of $A \cup B$ simplifies to the third axiom of probability.

If A and B are mutually exclusive events,

$$P(A \cup B) = P(A) + P(B) \tag{2-6}$$

Three or More Events

More complicated probabilities, such as $P(A \cup B \cup C)$, can be determined by repeated use of Equation 2-5 and by using some basic set operations. For example,

$$P(A \cup B \cup C) = P\left[(A \cup B) \cup C\right] = P(A \cup B) + P(C) - P\left[(A \cup B) \cap C\right]$$

Upon expanding $P(A \cup B)$ by Equation 2-5 and using the distributed rule for set operations to simplify $P\left[(A \cup B) \cap C\right]$, we obtain

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) - P(A \cap B) + P(C) - P\left[(A \cap C) \cup (B \cap C)\right] \\ &= P(A) + P(B) - P(A \cap B) + P(C) - \left[P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)\right] \end{aligned}$$

$= ( ) + ( ) + ( ) - \cap ( )\ B\ PA\ C\ PB\ C\ PA\ B\ C - \cap ( ) - \cap ( ) + \cap\cap ( )\ PA$

$PB\ PC\ PA$

We have developed a formula for the probability of the union of three events. Formulas can be developed for the probability of the union of any number of events, although the formulas become very complex. As a summary, for the case of three events,

$$PA\ B\ C\ PA\ PB\ PC\ PA\ B\ ( ) \cup \cup = ( ) + ( ) + ( ) - \cap ( )$$
$$- \cap\ PA\ C\ PB\ C\ PA\ B\ C\ ( ) - \cap ( ) + \cap\cap ( ) \quad (2\text{-}7)$$

Results for three or more events simplify considerably if the events are mutually exclusive. In general, a collection of events, $E\ ,E\ ,\ ,E\ ,\ _{1\ 2\ k} \dots$ is said to be mutually exclusive if there is no overlap among any of them. The Venn diagram for several mutually exclusive events is shown in Fig. 2-12. By generalizing the reasoning for the union of two events, the following result can be obtained:

E_1

E_3

E_4

E_2

FIGURE 2-12 Venn diagram of four mutually exclusive events.

(2-8)

Mutually Exclusive Events

A collection of events, $E\ ,E\ ,\ ,E\ ,\ _{1\ 2\ k} \dots$ is said to be **mutually exclusive** if for all pairs, $E_{i\ j} \cap = \varnothing\ E$

pH Here is a simple example of mutually exclusive events, which will be used quite frequently. Let X denote the pH of a sample. Consider the event that X is greater

to

7.8. This probability is the sum of any collection of mutually exclusive events with union equal to the same range for X. One example is

$$PX\ PX\ PX\ PX\ ( )\ 65\ 78\ 65\ 70\ 70\ 75\ 75\ 78\ .<\leq. = .\leq\leq. ( ) + ( ) .<\leq. + <\leq ( ) . .$$

Another example is

$$PX\ PX\ PX\ PX\ PX\ ( )\ 65\ 78\ 65\ 66\ 66\ 71\ 71\ 74\ 74\ .<\leq. = .<\leq. ( ) + .<\leq. ( ) + .<\leq. ( ) + .<\leq ( ) .\ 7\ 8$$

The best choice depends on the particular probabilities available.

Practical Interpretation: The partition of an event into mutually exclusive subsets is widely used in later chapters to calculate probabilities.

FOR SECTION 2-3

Problem available in *WileyPLUS* at instructor's discretion.

Tutoring problem available in *WileyPLUS* at instructor's discretion

2-82. If $P A( ) = . 0 3$, $PB , ( ) = . 0 2$ and $PA B , ( ) \cap =. 0 1$ deter mine the following probabilities:

(a) $P A( )'$ (b) $PA B ( ) \cup$ (c) $PA B (' \cap)$ (d) $P(A B \cap$ $')$ (e) $PA B [( )] \cup '$ (f) $PA B (' \cup)$

2-83. If A, B, and C are mutually exclusive events with $PA , ($ $) = . 0 2$ $PB , ( ) = . 0 3$ and $PC , ( ) = . 0 4$ determine the follow ing probabilities:

(a) $PA B C ( ) \cup \cup$ (b) $PA B C ( ) \cap \cap$ (c) $PA B ($ $) \cap$ (d) $PA B C [( )] \cup \cap$ (e) $PA B C ( ) ''' \cap \cap$

2-84. In the article "ACL Reconstruction Using Bone Patellar Tendon-Bone Press-Fit Fixation: 10-Year Clinical Results" in *Knee Surgery, Sports Traumatology, Arthroscopy* (2005, Vol. 13, pp. 248–255), the following causes for knee injuries were considered:

Activity	Percentage of Knee Injuries
Contact sport 46% Noncontact sport 44% Activity of daily living 9% Riding motorcycle 1%	

(a) What is the probability that a knee injury resulted from a sport (contact or noncontact)?

(b) What is the probability that a knee injury resulted from an activity other than a sport?

2-85. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

	Shock Resistance
	High Low
Scratch High 70 9 Resistance Low 16 5	

(a) If a disk is selected at random, what is the probability that its scratch resistance is high and its shock resistance is high?

(b) If a disk is selected at random, what is the probability that its

scratch resistance is high or its shock resistance is high?

(c) Consider the event that a disk has high scratch resistance and the event that a disk has high shock resistance. Are these two events mutually exclusive?

2-86. Strands of copper wire from a manufacturer are analyzed for strength and conductivity. The results from 100 strands are as follows:

	Strength
	High Low
High conductivity 74 8 Low conductivity 15 3	

Section 2-3/Addition Rules

(a) If a strand is randomly selected, what is the probability that

(a) $P A( )$ (b) $P B( )$

its conductivity is high and its strength is high?

(b) If a strand is randomly selected, what is the probability that its conductivity is low or its strength is low?

(c) Consider the event that a strand has low conductivity and the event that the strand has low strength. Are these two events mutually exclusive?

2-87. The analysis of shafts for a compressor is summa rized by conformance to specifications.

	Roundness Conforms
	Yes No
Surface Finish Yes 345 5 Conforms No 12 8	

(a) If a shaft is selected at random, what is the probability that it conforms to surface finish requirements?

(b) What is the probability that the selected shaft conforms to surface finish requirements or to roundness requirements? (c) What is the probability that the selected shaft either con forms to surface finish requirements or does not conform to roundness requirements?

(d) What is the probability that the selected shaft conforms to both surface finish and roundness requirements? 2-88. Cooking oil is produced in two main varieties: mono and polyunsaturated. Two common sources of cooking oil are corn and canola. The following table shows the number of bot tles of these oils at a supermarket:

	Type of oil
	Canola Corn
Type of Unsaturation Mono 7 13 Poly 93 77	

(a) If a bottle of oil is selected at random, what is the probabil ity that it belongs to the polyunsaturated category? (b) What is the probability that the chosen bottle is monoun saturated canola oil?

2-89. A manufacturer of front lights for automobiles tests lamps under a high-humidity, high-temperature environment using intensity and useful life as the responses of interest. The following table shows the performance of 130 lamps:

	Useful life
	Satisfactory Unsatisfactory
Intensity Satisfactory 117 3 Unsatisfactory 8 2	

(a) Find the probability that a randomly selected lamp will yield unsatisfactory results under any criteria.

(b) The customers for these lamps demand 95% satisfactory results. Can the lamp manufacturer meet this demand? 2-90. A computer system uses passwords that are six characters, and each character is one of the 26 letters (*a–z*) or 10 integers (0–9). Uppercase letters are not used. Let *A* denote the event that a password begins with a vowel (either *a*, *e*, *i*, *o*, or *u*), and let *B* denote the event that a password ends with an even number (either 0, 2, 4, 6, or 8). Suppose a hacker selects a password at random. Determine the following probabilities:

(c) *PA B* (∩) (d) *PA B* (∪)

2-91. Consider the endothermic reactions in Exercise 2-50. Let *A* denote the event that a reaction's final temperature is 271 K or less. Let *B* denote the event that the heat absorbed is above target. Use the addition rules to calculate the following probabilities.

(a) *PA B* (∪) (b) *PA B* (∩ ′) (c) *PA B* (′ ′ ∪) 2-92. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let *A* denote the event that the design color is red, and let *B* denote the event that the font size is not the smallest one. Use the addition rules to calculate the following probabilities. (a) *PA B* (∪) (b) *PA B* (∪ ′) (c) *PA B* (′ ′ ∪) 2-93. Consider the hospital emergency room data in Example 2-8. Let *A* denote the event that a visit is to hospital 4, and let *B* denote the event that a visit results in LWBS (at any hospital). Use the addition rules to calculate the following probabilities. (a) *PA B* (∪) (b) *PA B* (∪ ′) (c) *PA B* (′ ′ ∪) 2-94. Consider the well failure data in Exercise 2-53. Let *A* denote the event that the geological formation has more than 1000 wells, and let *B* denote the event that a well failed. Use the addition rules to calculate the following probabilities. (a) *PA B* (∪) (b) *PA B* (∪ ′) (c) *PA B* (′ ′ ∪) 2-95. Consider the bar code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Determine the probability for each of the following: (a) The first bar is wide or the second bar is wide. (b) Neither the first nor the second bar is wide.

(c) The first bar is wide or the second bar is not wide. (d) The first bar is wide or the first space is wide. 2-96. Consider the three patient groups in Exercise 2-57. Let *A* denote the event that the patient was treated with ribavirin plus interferon alfa, and let *B* denote the event that the response was complete. Determine the following probabilities: (a) *PA B* ( ) ∪ (b) *PA B* ( ) ′ ∪ (c) *PA B* ( ) ∪ ′ 2-97. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (*a–z*) or 26 uppercase letters (*A–Z*) or 10 integers (0–9). Assume all passwords are equally likely. Let *A* and *B* denote the events that consist of passwords with only letters or only integers, respectively. Determine the following probabilities: (a) *PA B* ( ) ∪ (b) *PA B* ( ) ′ ∪

(c) *P* (Password contains exactly 1 or 2 integers)

2-98. The article ["Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis," *Arthritis & Rheumatism* (2005, Vol. 52, pp. 3381– 3390)] considered four treatment groups. The groups consisted of patients with different drug therapies (such as prednisone and inf liximab): sequential monotherapy (group 1), step-up combination therapy (group 2), initial combination therapy (group 3), or ini tial combination therapy with infliximab (group 4). Radiographs of hands and feet were used to evaluate disease pro gression. The number of patients without progression of joint damage was 76 of

114 patients (67%), 82 of 112 patients (73%), 104 of 120 patients (87%), and 113 of 121 patients (93%) in groups 1–4, respectively. Suppose that a patient is selected randomly. Let *A* denote the event

2-4 Conditional Probability

that the patient is in group 1, and let *B* denote the event that there is no progression. Determine the following probabilities: (a) *PA B* ( ) ∪ (b) *PA B* ( ) ′ ′ ∪ (c) *PA B* ( ) ∪ ′

Sometimes probabilities need to be reevaluated as additional information becomes available. A useful way to incorporate additional information into a probability model is to assume that the out come that will be generated is a member of a given event. This event, say *A*, defi nes the conditions that the outcome is known to satisfy. Then probabilities can be revised to include this knowledge. The probability of an event B under the knowledge that the outcome will be in event A is denoted as $P(B|A)$

and this is called the **conditional probability of B given A**.

A digital communication channel has an error rate of 1 bit per every 1000 transmitted. Errors are rare, but when they occur, they tend to occur in bursts that affect many consecutive bits. If a single bit is transmitted, we might model the probability of an error as 1/1000. How ever, if the previous bit was in error because of the bursts, we might believe that the probability that the next bit will be in error is greater than 1/1000.

In a thin fi lm manufacturing process, the proportion of parts that are not acceptable is 2%. However, the process is sensitive to contamination problems that can increase the rate of parts that are not acceptable. If we knew that during a particular shift there were problems with the fi lters used to control contamination, we would assess the

In a manufacturing process, 10% of the parts contain visible surface fl aws and 25% of the parts with surface fl aws are (functionally) defective parts. However, only 5% of parts without surface fl aws are defective parts. The probability of a defective part depends on our knowledge of the pres ence or absence of a surface fl aw. Let D denote the event that a part is defective, and let F denote the event that a part has a surface fl aw. Then we denote the probability of D given or assuming that a part has a surface fl aw, as $P(D|F)$. Because 25% of the parts with surface fl aws are defective, our conclusion can be stated as $P(D|F) = 0.25$. Furthermore, because F' denotes the event that a part does not have a surface fl aw and because 5% of the parts without surface fl aws are defective, we have $P(D|F') = 0.05$. These results are shown graphically in Fig. 2-13.

Surface Flaws and Defectives Table 2-3 provides an example of 400 parts classifi ed by surface fl aws and as

previously in this section. For example, of the parts with surface fl aws (40 parts), the number of defective ones is 10.

Therefore, $P(D|F) = 10/40 = 0.25$

and of the parts without surface fl aws (360 parts), the number of defective ones is 18.

Therefore, $P(D|F') = 18/360 = 0.05$

Practical Interpretation: The probability of being defective is fi ve times greater for parts with surface fl aws. This calculation illustrates how probabilities are adjusted for additional information. The result also suggests that there may be a link between surface fl aws and functionally defective parts, which should be investigated.

2-3 Parts Classifi ed

		Surface Flaws		
		Yes (event F)	No	Total
Defective	Yes (event D)	10	18	28
	No	30	342	372
	Total	40	360	400

P(DuF) = 0.25

25%
defective

5% defective

P(DuF′) = 0.05

FIGURE 2-13 Conditional probabilities for parts with surface fl aws.

F = parts with surface flaws

F′ = parts without surface flaws

PA B PA $^{A\ B}$ ( ) () ∩ = ∩

/ number of outcomes in

Conditional Probability

In Example 2-22, conditional probabilities were calculated directly. These probabilities can also be determined from the formal defi nition of conditional probability.

The **conditional probability** of an event *B* given an event *A*, denoted as *PBA* ( ) | , is *PBA PA B PA* ( ) | /

= ∩ ( ) () (2-9)

for *P A*( ) > 0.

This defi nition can be understood in a special case in which all outcomes of a random experiment are equally likely. If there are *n* total outcomes,

PA A n ( ) = ( ) number of outcomes in /

Also,

PA B A B ( ) ∩ = ∩ ( ) number of outcomes in / *n*

Consequently,

number of outcomes in *A*

Therefore, *PBA* ( ) | can be interpreted as the relative frequency of event *B* among the trials that produce an outcome in event *A*.

in Table 2-3. From this table, *PDF PD F PF* ( )

| = ∩ ( ) () = = 10

Tree Diagram Again consider the 400 parts

40

10

400

400

40

ties are

=

/ |/

40 400 10 28

( ) = ( ) =

PD PDF

28 400 10 40

/ l/

Here, *P D*( ) and *PDF* ( ) l are probabilities of the same event, but they are computed under two different states of knowledge. Similarly, *P F*( ) and *PF D* ( ) l are computed under two different states of knowledge. The tree diagram in Fig. 2-14 can also be used to display conditional probabilities. The fi rst branch is on surface fl aw. Of the 40 parts with surface fl aws, 10 are functionally defective and 30 are not. Therefore,

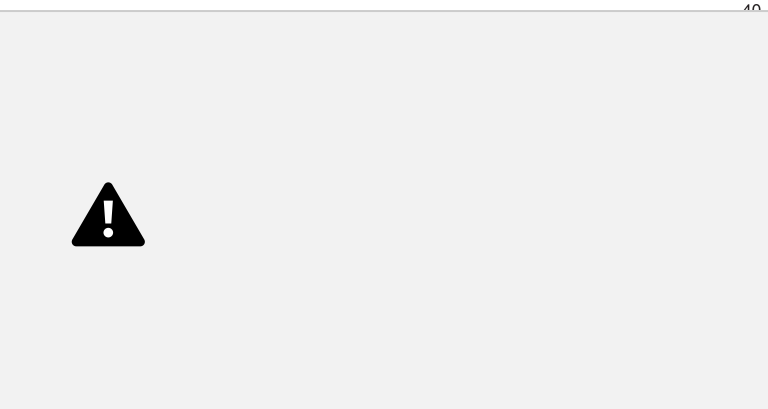
PD F PD F ( ) l / (l) / = = 10 40 30 40 and ′

Of the 360 parts without surface fl aws, 18 are functionally defective and 342 are not.

Therefore, *PDF PD F* ( ) l/ l / ′ ″ = 18 360 342 360 and ( ) =

Surface flaw

40



No Yes

360

40

40

FIGURE 2-14 Tree diagram for parts classifi ed.

Random Samples

Random Samples and Conditional Probability

Recall that to select one item randomly from a batch implies that each item is equally likely to be picked. If more than one item is selected, *randomly* implies that each element of the sample space is equally likely to be picked. When sample spaces were presented earlier in this chap ter, sampling with and without replacement was defi ned and illustrated for the simple case of a batch with three items {*abc* ,,}. If two items are selected randomly from this batch without replacement, each of the six outcomes in the ordered sample space { } *ab, ac,ba,bc,ca,cb* has probability 1 6/ . If the unordered sample space is used, each of the three outcomes in {{ , }, *a b* {*ac bc* , },{ , }} has probability 1 3/ .

When a sample is selected randomly from a large batch, it is usually easier to avoid enu meration of the sample space and calculate probabilities from conditional probabilities. For example, suppose that a batch contains 10 parts from tool 1 and 40 parts from tool 2. If two parts are selected randomly, without replacement, what is the conditional probability that a part from tool 2 is selected second given that a part from tool 1 is selected fi rst?

Although the answer can be determined from counts of outcomes, this type of question can be answered more easily with the following result.

To select *randomly* implies that at each step of the sample, the items that remain in the batch are equally likely to be selected.

If a part from tool 1 were selected with the fi rst pick, 49 items would remain, 9 from tool 1 and 40 from tool 2, and they would be equally likely to be picked. Therefore, the probability that a part from tool 2 would be selected with the second pick given this fi rst pick is

PE E ( ) $_{2\ 1}$ l / = 40 49

In this manner, other probabilities can also be simplifi ed. For example, let the event *E* consist of the outcomes with the first selected part from tool 1 and the second part from tool 2. To determine the probability of E, consider each step. The probability that a part from tool 1 is selected with the first pick is $P(E_1) = 10/50$. The conditional probability that a part from tool 2 is selected with the second pick, given that a part from tool 1 is selected first, is $P(E_2 | E_1) = 40/49$. Therefore,

$$P(E) = P(E_2 | E_1)P(E_1) = \frac{40}{49} \cdot \frac{10}{50} = \frac{8}{49}$$

Sometimes a partition of the question into successive picks is an easier method to solve the problem.

Random Inspection Consider the inspection described in Example 2-14. Six parts are selected randomly without replacement from a bin of 50 parts. The bin contains 3 defective parts and 47 nondefective parts. What is the probability that the second part is defective given that the first part is defective?

Let A and B denote the events that the first and second part selected are defective, respectively. The probability requested can be expressed as $P(B | A)$. If the first part is defective, prior to selecting the second part the batch contains 49 parts, of which 2 are defective. Therefore,

$$P(B | A) = \frac{2}{49}$$

Continuing Example 2-24, what is the probability that the first two parts selected are defective and the third is not defective?

This probability can be described in shorthand notation as

that are defective
and not defective,
respectively. Here
P ddn P n dd P dd
P n dd P d d P d ( )
(|)( ) (|)(|)( ) 123 3
47 2 =
12 12 3 12 2 1 1 48
= =· 3
49 50

· = 0 0024 .

The probabilities for the first and second selections are similar to those in the previous example. The $P(n_3|d_1d_2)$ is based on the fact that after the first 2 parts are selected, 1 defective and 47 nondefective parts remain. When the probability is written to account for the order of the selections, it is easy to solve this question from the definition of conditional probability. There are other ways to express the probability, such as $P(d_1d_2n_3) = P(d_2|d_1n_3)\,P(d_1n_3)$. However, such alternatives do not lead to conditional probabilities that can be easily calculated.

FOR SECTION 2-4

Problem available in *WileyPLUS* at instructor's discretion.
Tutoring problem available in *WileyPLUS* at instructor's discretion

2-99. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch Resistance	High	70	9
	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$
(c) $P(A|B)$ (d) $P(B|A)$

2-100. Samples of skin experiencing desquamation are analyzed for both moisture and melanin content. The results from 100 skin samples are as follows:

		Melanin Content	
		High	Low
Moisture Content	High	13	7
	Low	48	32

Let A denote the event that a sample has low melanin content, and let B denote the event that a sample has high moisture content. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$
(c) $P(A|B)$ (d) $P(B|A)$

2-101. The analysis of results from a leaf transmutation experiment (turning a leaf into a petal) is summarized by type of transformation completed:

		Total Textural Transformation	
		Yes	No
Total Color Transformation	Yes	243	26
	No	13	18

(a) If a leaf completes the color transformation, what is the probability that it will complete the textural transformation?
(b) If a leaf does not complete the textural transformation, what is the probability it will complete the color transformation?

2-102. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and length measurements. The results of 100 parts are summarized as follows:

		Length	
		Excellent	Good
Surface Finish	Excellent	80	2
	Good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent length. Determine:

(a) $P(A)$ (b) $P(B)$
(c) $P(A|B)$ (d) $P(B|A)$

(e) If the selected part has excellent surface finish, what is the probability that the length is excellent?
(f) If the selected part has good length, what is the probability that the surface finish is excellent?

2-103. The following table summarizes the analysis of sam ples of galvanized steel for coating weight and surface roughness:

		Coating Weight	
		High	Low
Surface Roughness	High	12	16
	Low	88	34

(a) If the coating weight of a sample is high, what is the probability that the surface roughness is high?

(b) If the surface roughness of a sample is high, what is the probability that the coating weight is high?

(c) If the surface roughness of a sample is low, what is the probability that the coating weight is low?

2-104. Consider the data on wafer contamination and location in the sputtering tool shown in Table 2-2. Assume that one wafer is selected at random from this set. Let A denote the event that a wafer contains four or more particles, and let B denote the event that a wafer is from the center of the sputtering tool. Determine:

(a) $P(A)$ (b) $P(A \mid B)$

(c) $P(B)$ (d) $P(B \mid A)$

(e) $P(A \cap B)$ (f) $P(A \cup B)$

2-105. The following table summarizes the number of deceased beetles under autolysis (the destruction of a cell after its death by the action of its own enzymes) and putrefaction (decomposition of organic matter, especially protein, by micro organisms, resulting in production of foul-smelling matter):

		Autolysis	
		High	Low
Putrefaction	High	14	59
	Low	18	9

(a) If the autolysis of a sample is high, what is the probability that the putrefaction is low?

(b) If the putrefaction of a sample is high, what is the probability that the autolysis is high?

(c) If the putrefaction of a sample is low, what is the probability that the autolysis is low?

2-106. A maintenance firm has gathered the following information regarding the failure mechanisms for air condi tioning systems:

		Evidence of Gas Leaks	
		Yes	No
Evidence of electrical failure	Yes	55	17
	No	32	3

The units without evidence of gas leaks or electrical failure showed other types of failure. If this is a representative sample of AC failure, find the probability

(a) That failure involves a gas leak

(b) That there is evidence of electrical failure given that there was a gas leak

(c) That there is evidence of a gas leak given that there is evidence of electrical failure

2-107. A lot of 100 semiconductor chips contains 20 that are defective. Two are selected randomly, without replacement, from the lot.

(a) What is the probability that the first one selected is defective?

(b) What is the probability that the second one selected is defective given that the first one was defective?

(c) What is the probability that both are defective? (d) How does the answer to part (b) change if chips selected were replaced prior to the next selection?

2-108. A batch of 500 containers for frozen orange juice contains 5 that are defective. Two are selected, at random, with out replacement from the batch.

(a) What is the probability that the second one selected is defective given that the first one was defective?

(b) What is the probability that both are defective? (c) What is the probability that both are acceptable? Three containers are selected, at random, without replace ment, from the batch.

(d) What is the probability that the third one selected is defective given that the first and second ones selected were defective? (e) What is the probability that the third one selected is defective given that the first one selected was defective and the second one selected was okay?

(f) What is the probability that all three are defective?

2-109. A batch of 350 samples of rejuvenated mitochondria contains 8 that are mutated (or defective). Two are selected from the batch, at random, without replacement. (a) What is the probability that the second one selected is defective given that the first one was defective?

(b) What is the probability that both are defective? (c) What is the probability that both are acceptable?

2-110. A computer system uses passwords that are exactly seven characters and each character is one of the 26 letters (*a–z*) or 10 integers (0–9). You maintain a password for this computer system. Let A denote the subset of passwords that begin with a vowel (either *a*, *e*, *i*, *o*, or *u*) and let B denote the subset of passwords that end with an even number (either 0, 2, 4, 6, or 8). (a) Suppose a hacker selects a password at random. What is the probability that your password is selected?

(b) Suppose a hacker knows that your password is in event A and selects a password at random from this subset. What is the probability that your password is selected?

(c) Suppose a hacker knows that your password is in A and B and selects a password at random from this subset. What is the probability that your password is selected?

2-111. If $P(A \mid B) = 1$, must $A = B$? Draw a Venn diagram to explain your answer.

2-112. Suppose A and B are mutually exclusive events. Construct a Venn diagram that contains the three events A, , B and C such that $P(A \mid C) = 1$ and $P(B \mid C) = 0$.

2-113. Consider the endothermic reactions in Exercise 2-50. Let A denote the event that a reaction's fi nal temperature is 271 K or less. Let B denote the event that the heat absorbed is above target. Determine the following probabilities.

(a) $P(A|B)$ (b) $P(A'|B)$

(c) $P(A|B')$ (d) $P(B|A)$

2-114. Consider the hospital emergency room data in Exam ple 2-8. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hos pital). Determine the following probabilities.

(a) $P(A|B)$ (b) $P(A'|B)$

(c) $P(A|B')$ (d) $P(B|A)$

2-115. Consider the well failure data in Exercise 2-53. (a) What is the probability of a failure given there are more than 1,000 wells in a geological formation?

(b) What is the probability of a failure given there are fewer than 500 wells in a geological formation?

2-116. An article in the *The Canadian Entomologist* (Har court et al., 1977, Vol. 109, pp. 1521–1534) reported on the life of the alfalfa weevil from eggs to adulthood. The follow ing table shows the number of larvae that survived at each stage of development from eggs to adults.

Eggs	Early Larvae	Late Larvae	Pre pupae	Late Pupae	Adults
421	412	306	45	35	31

(a) What is the probability an egg survives to adulthood? (b) What is the probability of survival to adulthood given survival to the late larvae stage?

(c) What stage has the lowest probability of survival to the next stage?

2-117. Consider the bar code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Determine the probability for each of the following: (a) The second bar is wide given that the fi rst bar is wide. (b) The third bar is wide given that the fi rst two bars are not wide. (c) The fi rst bar is wide given that the last bar is wide.

2-118. Suppose that a patient is selected randomly from those described in Exercise 2-57. Let A denote the event that the patient is treated with ribavirin plus interferon alfa, and let B denote the event that the response is complete. Determine the following probabilities:

(a) $P(B|A)$ (b) $P(A|B)$

(c) $P(A|B')$ (d) $P(A'|B)$

2-119. Suppose that a patient is selected randomly from those described in Exer cise 2-98. Let A denote the event that the patient is in group 1, and let B denote the event that there is no progres sion. Determine the following probabilities:

(a) $P(B|A)$ (b) $P(A|B)$

(c) $P(A|B')$ (d) $P(A'|B)$

2-120. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (*a–z*) or 26 uppercase letters (*A–Z*) or 10 integers (0–9). Let Ω denote the set of all possible passwords. Suppose that all passwords in Ω are equally likely. Determine the probability for each of the following:

(a) Password contains all lowercase letters given that it con tains only letters

(b) Password contains at least 1 uppercase letter given that it contains only letters

(c) Password contains only even numbers given that is con tains all numbers

2-5 Multiplication and Total Probability Rules

The probability of the intersection of two events is often needed. The conditional probability defi - nition in Equation 2-9 can be rewritten to provide a formula known as the **multiplication rule** for probabilities.

Multiplication Rule

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B) \quad (2\text{-}10)$$

The last expression in Equation 2-10 is obtained by interchanging A and B.

of a numerically controlled machining operation for

Failures are

fi xture alignment, cutting blade condition, vibration, and ambient environmental conditions. Given that the first stage meets specifications, the probability that a second stage of machining meets specifications is 0.95. What is the probability that both stages meet specifications?

Let A and B denote the events that the first and second stages meet specifications, respectively. The probability requested is

$$P(A \cap B) = P(B \mid A)P(A) = 0.95(0.90) = 0.855$$

Although it is also true that $P(A \cap B) = P(A \mid B)P(B)$, the information provided in the problem does not match this second formulation.

Practical Interpretation: The probability that both stages meet specifications is approximately 0.85, and if additional stages were needed to complete a piston, the probability would decrease further. Consequently, the probability that each stage is completed successfully needs to be large in order for a piston to meet all specifications.

Sometimes the probability of an event is given under each of several conditions. With enough of these conditional probabilities, the probability of the event can be recovered. For example, suppose that in semiconductor manufacturing, the probability is 0.10 that a chip subjected to high levels of contamination during manufacturing causes a product failure. The probability is 0.005 that a chip not subjected to high contamination levels during manufacturing causes a product failure. In a particular production run, 20% of the chips are subject to high levels of contamination. What is the probability that a product using one of these chips fails?

Clearly, the requested probability depends on whether or not the chip was exposed to high levels of contamination. For any event B, we can write B as the union of the part of B in A and the part of B in A'. That is,

$$B = (A \cap B) \cup (A' \cap B)$$

This result is shown in the Venn diagram in Fig. 2-15. Because A and A' are mutually exclusive, $A \cap B$ and $A' \cap B$ are mutually exclusive. Therefore, from the probability of the union of mutually exclusive events in Equation 2-6 and the multiplication rule in Equation 2-10, the following **total probability rule** is obtained.

Total Probability Rule (Two Events)

For any events A and B,

$$P(B) = P(B \cap A) + P(B \cap A') = P(B \mid A)P(A) + P(B \mid A')P(A') \qquad (2\text{-}11)$$

Probability of Failure	Level of Contamination	
		0.1 Hi
0.005 Not high 0.8		

Semiconductor Contamination Consider the contamination discussion at the start of this section. The information is summarized here.

Let F denote the event that the product fails, and let H denote the event that the chip is exposed to high levels of con tamination. The requested probability is $P F( )$, and the information provided can be represented as $( ) \mid\mid = . ( ) = .$

$P F H PF H$

0 10 0 005

and ′

$( ) =. ( ) = .$ 0 20 0 80 $PH PH$

and ′

From Equation 2-11,

$P F( ) =. . 0\ 10\ 0\ 20\ 0\ 005\ 0\ 80\ 0\ 024 ( ) +. . ( ) = .$

which can be interpreted as just the weighted average of the two probabilities of failure.

The reasoning used to develop Equation 2-11 can be applied more generally. Because $A \cup$ $' = A\ S$, we know $(AB\ A\ B \cap) \cup \cap ( )'$ equals B, and because $A \cap ' = \varphi$, A we know $A \cap B$ and $A' \cap B$ are mutually exclusive. In general, a collection of sets $E_{1\,2}, E, , E_k \ldots$ such that $E_{1\,2} \cup$ $\cup \ldots \cup = E\ ES_{\,k}$ is said to be **exhaustive**. A graphical display of partitioning an event B among a collection of mutually exclusive and exhaustive events is shown in Fig. 2-16.

Total Probability Rule (Multiple Events)

Assume $E_{1\,2}, E, , E_k \ldots$ are k mutually exclusive and exhaustive sets. Then $\cdots +$

$( ) = \cap ( ) + \cap ( ) + \cap ( )_k$

$PB\ PB\ E\ PB\ E\ PB\ E$

1 2

$\mid\mid \cdots + PBE\ PE_{\,k\,k} ( )( ) \mid$ (2-12)

$= ( )( ) + ( )( ) +$

$PBE\ PE\ PBE\ PE$

11 2 2

$A\ A'$

$B > A'$

$B > A$

B

FIGURE 2-15 Partitioning an event into two mutually exclusive subsets.

E_1

$B > E_1$

2 3 2

$E \quad E \quad E_4\ B > E \quad B > E_3$

$B > E_4$

FIGURE 2-16 Partitioning an event into several mutually exclusive subsets.

$B = (B > E_1) < (B > E_2) < (B > E_3) < (B > E_4)$

Probability of Failure	Level of C
	High
	Medium
	Low

In a particular production run, 20% of the chips are subjected to high levels of contamination, 30% to medium levels of contamination, and 50% to low levels of contamination. What is the probability that a product using one of these chips fails? Let

H denote the event that a chip is exposed to high levels of contamination
M denote the event that a chip is exposed to medium levels of contamination
L denote the event that a chip is exposed to low levels of contamination

Then,

$$PF\ PF\ HPH\ PF\ MPM\ PF\ LPL\ (\ ) = (\ )(\ ) + (\ )(\ ) + (\ )(\ )$$
$$| |\ |$$

$$=.\ .\ (\ ) +.\ .\ ($$
$$0\ 10\ 0\ 20\ 0\ 01\ 0\ 30) +.\ .\ 0\ 001\ 0\ 50\ 0\ 0235\ (\ ) = .$$

The calculations are conveniently organized with the tree diagram in Fig. 2-17.

Contamination

0.20 0.50

0.30

High Medium Low

0.90

P(FailuMedium) 5 0.01

FailuMedium) 5 0.99

0.001

0.999(0.50) 5 0.4995

P(FailuHigh) 5 0.10

0.001(0.50) 5 0.0005

0.01(0.30) 5 0.003

0.99(0.30) 5 0.297

0.10(0.20) 5 0.02

0.90(0.20) 5 0.18

P(Not FailuLow) 5 0.999

P(Not FailuHigh) 5

P(Not

P(FailuLow) 5

P(Fail) 5 0.02 + 0.003 + 0.0005 5 0.0235

FIGURE 2-17 Tree diagram for Example 2-28.

FOR SECTION 2-5

Problem available in *WileyPLUS* at instructor's discretion.
Tutoring problem available in *WileyPLUS* at instructor's discretion

2-121. Suppose that $PAB\ (\ |\ ) = .\ 0\ 4$ and $P\ B(\ ) = ..\ 0\ 5$ Deter mine the following:
(a) $PA\ B\ (\ \cap\ )$ (b) $PA\ B\ (\ '\cap\ )$

2-122. Suppose that $P\ A\ B\ ,P\ A\ B\ ,\ (\ |\ |\ ) =.\ 02\ 03\ (\ '\ ) = .$ and $P\ B(\ ) = ..\ 0\ 8$ What is $P\ A(\ )$?

2-123. The probability is 1% that an electrical connector that is kept dry fails during the warranty period of a portable computer. If the connector is ever wet, the probability of a fail ure during the warranty period is 5%. If 90% of the connectors are kept dry and 10% are wet, what proportion of connectors fail during the warranty period?

2-124. Suppose 2% of cotton fabric rolls and 3% of nylon fabric rolls contain flaws. Of the rolls used by a manufacturer, 70% are cotton and 30% are nylon. What is the probability that a randomly selected roll used by the manufacturer contains flaws?

2-125. The edge roughness of slit paper products increases as knife blades wear. Only 1% of products slit with new blades have rough edges, 3% of products slit with blades of average sharpness exhibit roughness, and 5% of products slit with worn blades exhibit roughness. If 25% of the blades in manufacturing are new, 60% are of average sharpness, and 15% are worn, what is the proportion of products that exhibit edge roughness?

2-126. In the 2012 presidential election, exit polls from the critical state of Ohio provided the following results:

Total Obama Romney
No college degree (60%) 52% 45% College
graduate (40%) 47% 51%

What is the probability a randomly selected respondent voted for Obama?

2-127. Computer keyboard failures are due to faulty electrical connects (12%) or mechanical defects (88%). Mechanical defects are related to loose keys (27%) or improper assembly (73%). Electrical connect defects are caused by defective wires (35%), improper connections (13%), or poorly welded wires (52%).
(a) Find the probability that a failure is due to loose keys. (b) Find the probability that a failure is due to improperly connected or poorly welded wires.

2-128. Heart failures are due to either natural occurrences

(87%) or outside factors (13%). Outside factors are related to induced substances (73%) or foreign objects (27%). Natural occurrences are caused by arterial blockage (56%), disease (27%), and infection (e.g., staph infection) (17%).

(a) Determine the probability that a failure is due to an induced substance.

(b) Determine the probability that a failure is due to disease or infection.

2-129. A batch of 25 injection-molded parts contains 5 parts that have suffered excessive shrinkage.

(a) If two parts are selected at random, and without replace ment, what is the probability that the second part selected is one with excessive shrinkage?

(b) If three parts are selected at random, and without replace ment, what is the probability that the third part selected is one with excessive shrinkage?

2-130. A lot of 100 semiconductor chips contains 20 that are defective.

(a) Two are selected, at random, without replacement, from the lot. Determine the probability that the second chip selected is defective.

(b) Three are selected, at random, without replacement, from the lot. Determine the probability that all are defective. 2-131. An article in the *British Medical Journal* ["Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy" (1986, Vol. 82, pp. 879–892)] provided the following discussion of success rates in kidney stone removals. Open surgery had a success rate of 78% (273/350) and a newer method, percutane ous nephrolithotomy (PN), had a success rate of 83% (289/350). This newer method looked better, but the results changed when stone diameter was considered. For stones with diameters less than 2 centimeters, 93% (81/87) of cases of open surgery were successful compared with only 83% (234/270) of cases of PN. For stones greater than or equal to 2 centimeters, the success rates were 73% (192/263) and 69% (55/80) for open surgery and PN, respectively. Open surgery is better for both stone sizes, but less successful in total. In 1951, E. H. Simpson pointed out this apparent contradiction (known as **Simpson's paradox**), and the hazard still persists today. Explain how open surgery can be bet ter for both stone sizes but worse in total.

2-132. Consider the endothermic reactions in Exercise 2-50. Let *A* denote the event that a reaction's final temperature is 271 K or less. Let *B* denote the event that the heat absorbed is above target. Determine the following probabilities.

(a) *PA B* (∩) (b) *PA B* (∪) (c) *PA B* (′ ′ ∪) (d) Use the total probability rule to determine *P A*( )

2-133. Consider the hospital emergency room data in Exam ple 2-8. Let *A* denote the event that a visit is to hospital 4 and let *B* denote the event that a visit results in LWBS (at any hos pital). Determine the following probabilities.

(a) *PA B* (∩) (b) *PA B* (∪) (c) *PA B* ″ ′ (∪) (d) Use the total probability rule to determine *P A*( )

2-134. Consider the hospital emergency room data in Example 2-8. Suppose that three visits that resulted in LWBS are selected randomly (without replacement) for a follow-up interview. (a) What is the probability that all three are selected from hospital 2?

(b) What is the probability that all three are from the same hospital?

2-135. Consider the well failure data in Exercise 2-53. Let *A* denote the event that the geological formation has more than 1000 wells, and let *B* denote the event that a well failed. Deter mine the following probabilities.

(a) *PA B* (∩) (b) *PA B* (∪) (c) *PA B* (′ ′ ∪) (d) Use the total probability rule to determine *P*( ) *A* 2-136. Consider the well failure data in Exercise 2-53. Sup pose that two failed wells are selected randomly (without replacement) for a follow-up review.

(a) What is the probability that both are from the gneiss geo logical formation group?

(b) What is the probability that both are from the same geo logical formation group?

2-137. A Web ad can be designed from four different colors, three font types, fi ve font sizes, three images, and fi ve text phrases. A specifi c design is randomly generated by the Web server when you visit the site. Determine the probability that the ad color is red and the font size is not the smallest one.

2-138. Consider the code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Determine the probability for each of the following: (a) The code starts and ends with a wide bar.

(b) Two wide bars occur consecutively.

(c) Two consecutive wide bars occur at the start or end. (d) The middle bar is wide.

2-139. Similar to the hospital schedule in Example 2-11, sup pose that an oper ating room needs to schedule three knee, four

2-6 Independence

hip, and fi ve shoulder surgeries. Assume that all schedules are equally likely. Determine the following probabil ities: (a) All hip surgeries are completed fi rst given that all knee surgeries are last.

(b) The schedule begins with a hip surgery given that all knee surgeries are last.

(c) The fi rst and last surgeries are hip surgeries given that knee surgeries are scheduled in time periods 2 through 4. (d) The fi rst two surgeries are hip surgeries given that all knee surgeries are last.

2-140. Suppose that a patient is selected randomly from those described in Exercise 2-98. Let *A* denote the event that the patient is in group 1, and let *B* denote the event for which there is no progression. Determine the following probabilities:

(a) *PA B* ( ) ∩ (b) *P B*( )

(c) *PA B* ( ) ′ ∩ (d) *PA B* ( ) ∪ (e) *PA B* ( ) ′ ∪ 2-141. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (*a–z*) or 26 uppercase letters (*A–Z*) or 10 integers (0–9). Let Ω denote the set of all possible password, and let *A* and *B* denote the events that consist of passwords with only letters or only integers, respectively. Suppose that all passwords in Ω are equally likely. Determine the following

robabilities: (a) $P(A|B')$
(b) $P(A' \cap B)$
(c) P (password contains exactly 2 integers given that it con tains at least 1 integer)

might equal $P(B)$. In this special case, knowledge that the outcome of the experiment is in event A does not affect the probability that the outcome is in event B.

Sampling with Replacement Consider the inspection described in Example 2-14. Six parts are selected ran domly from a bin of 50 parts, but assume that the selected part is replaced before the next

one is selected. The bin contains 3 defective parts and 47

nondefective parts. What is the probability that the second part is defective given that the fi rst part is defective?

In shorthand notation, the requested probability is $P(B \mid A)$, where A and B denote the events that the fi rst and second parts are defective, respectively. Because the fi rst part is replaced prior to selecting the second part, the bin still contains 50 parts, of which 3 are defective. Therefore, the probability of B does not depend on whether or not the fi rst part is defective. That is,

$$P(B|A) = \frac{3}{50}$$

Also, the probability that both parts are defective is

$$P(A \cap B) = P(B|A)P(A) = \frac{3}{50} \cdot \frac{3}{50} = \frac{9}{2500}$$

that case, we determined that $P(D|F) = 10/40 = 0.25$

Suppose that the situation is different and follows Table 2-4. Then,

$$P(D|F) = 2/40 = 0.05 \quad \text{and} \quad P(D) = 20/400 = 0.05$$

That is, the probability that the part is defective does not depend on whether it has surface fl aws.

Also, $P(F|D) = 2/20 = 0.10$ and $P(F) = 40/400 = 0.10$

so the

probability of a surface flaw does not depend on whether the part is defective. Furthermore, the definition of conditional probability implies that

$$P(F \cap D) = P(D \mid F)P(F)$$

but in the special case of this problem,

$$P(F \cap D) = P(D)P(F) = \frac{2}{40}\frac{2}{20} = \frac{1}{200}.$$

2-4 Parts Classified

		Surface Flaws		
		Yes (event F)	No	Total
Defective	Yes (event D)	2	18	20
	No	38	342	380
	Total	40	360	400

The preceding example illustrates the following conclusions. In the special case that $P(B \mid A) = P(B)$, we obtain

$$P(A \cap B) = P(B \mid A)P(A) = P(B)P(A)$$

and

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

These conclusions lead to an important definition.

Independence (two events)

Two events are **independent** if any one of the following equivalent statements is true:

(1) $P(A \mid B) = P(A)$

(2) $P(B \mid A) = P(B)$

(3) $P(A \cap B) = P(A)P(B)$ (2-13)

It is left as a mind-expanding exercise to show that

relationship between events and is used throughout this text. A mutually exclusive relationship between two events is based only on the outcomes that compose the events. However, an independence relationship depends on the probability model used for the random experiment. Often, independence is assumed to be part of the random experiment that describes the physical system under study.

Consider the inspection described in Example 2-14. Six parts are selected randomly without replacement from a bin of 50 parts. The bin contains 3 defective

We suspect that these two events are not independent because the knowledge that the first part is defective suggests that it is less likely that the second part selected is defective. Indeed, $P(B \mid A) = 2/49$. Now, what is $P(B)$? Finding the unconditional $P(B)$ takes some work because the possible values of the first selection need to be considered:

$$P(B) = P(B \mid A)P(A) + P(B \mid A')P(A')$$

$$\frac{2}{49} \cdot \frac{3}{50} + \frac{3}{49} \cdot \frac{47}{50} = \frac{3}{50}$$

Interestingly, $P(B)$, the unconditional probability that the second part se lected is defective, without any knowl edge of the fi rst part, is the same as the probability that the fi rst part selected is defective. Yet our goal is to assess independence. Because $P(B \mid A)$ does not equal $P(B)$, the two events are not independent, as we expected.

When considering three or more events, we can extend the defi nition of independence with the following general result.

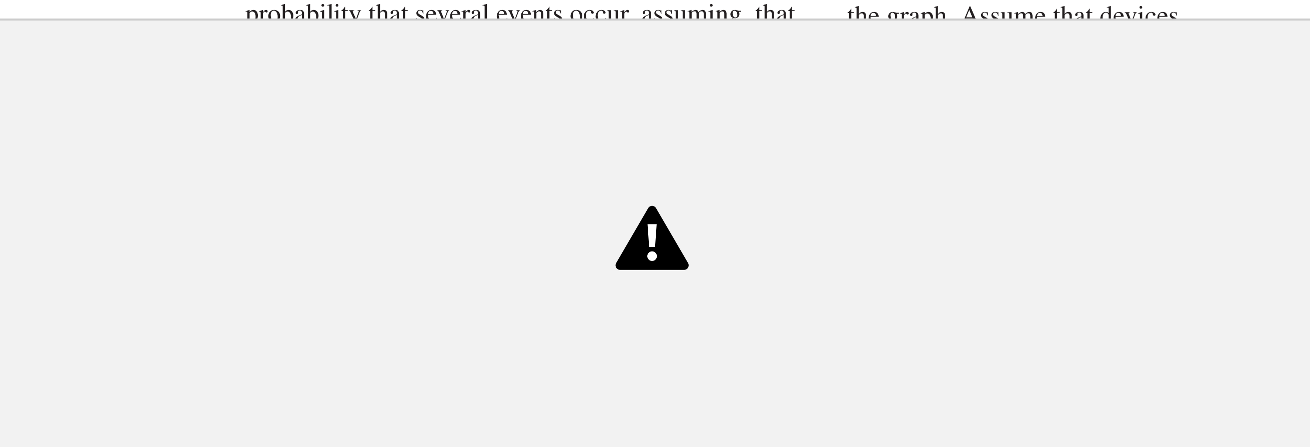
Independence (multiple events)

The events $E_1, E_2, \ldots, E_n$ are independent if and only if for any subset of these events $E_{i_1}, E_{i_2}, \ldots, E_{i_k}$

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = P(E_{i_1}) \cdot P(E_{i_2}) \cdots P(E_{i_k}) \quad (2\text{-}14)$$

This defi nition is typically used to calculate the probability that several events occur, assuming that they are independent usually comes from a fundamental understanding of the random experiment.

Series Circuit The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



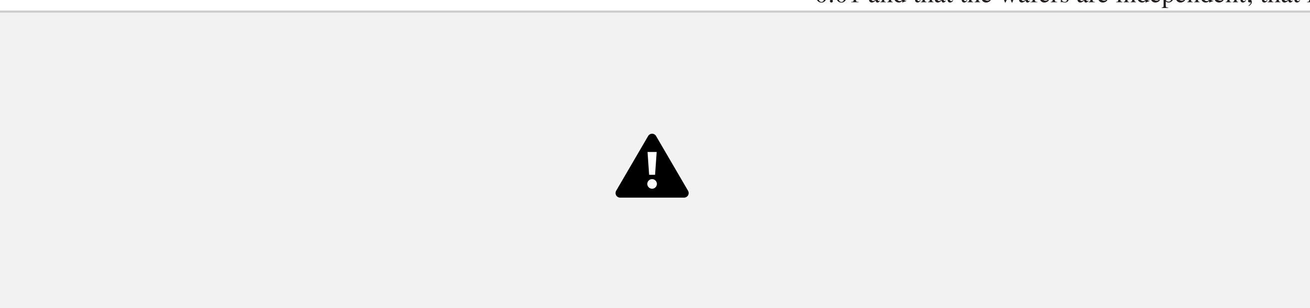
0.8 0.9

Let L and R denote the events that the left and right devices operate, respectively. There is a path only if both oper ate. The probability that the circuit operates is

$$P(L \text{ and } R) = P(L \cap R) = P(L)P(R) = 0.80(0.90) = 0.72$$

Practical Interpretation: Notice that the probability that the circuit operates degrades to approximately 0.5 when all devices are required to be functional. The probability that each device is functional needs to be large for a circuit to operate when many devices are connected in series.

that a wafer contains a large particle of con tamination is 0.01 and that the wafers are independent; that is, the



a large particle does not depend on the

characteristics of any of the other wafers. If 15 wafers are analyzed, what is the probability that no large particles are found?

Let E_i denote the event that the ith wafer contains no large particles, $i = 1, 2, \ldots, 15$. Then, $P(E_i) = 0.99$. The probability requested can be represented as $P(E_1 \cap E_2 \cap \cdots \cap E_{15})$. From the independence assumption and Equation 2-14, $P(E_1 \cap E_2 \cap \cdots \cap E_{15}) = P(E_1) \cdot P(E_2) \cdots P(E_{15}) = 0.99^{15} = 0.86$

there is a path of functional devices from left to right. The probability that each device functions is shown on the

fail independently. What is the probability that the circuit operates?

0.95

a b

0.95

Let T and B denote the events that the top and bottom devices operate, respectively. There is a path if at least one device operates. The probability that the circuit operates is

$$P(T \text{ or } B) = 1 - P[(T \text{ or } B)'] = 1 - P(T' \text{ and } B')$$

A simple formula for the solution can be derived from the complements T' and B'. From the independence assumption,

$$P(T' \text{ and } B') = P(T')P(B') = (1 - 0.95)^2 = 0.05^2$$

so

$$P(T \text{ or } B) = 1 - 0.05^2 = 0.9975$$

Practical Interpretation: Notice that the probability that the circuit operates is larger than the probability that either device is functional. This is an advantage of a parallel architecture. A disadvantage is that multiple devices are needed.

there is a path of functional devices from left to right. The probability that each device functions is shown on the

devices fail independently. What is the probability that the circuit operates?

0.9

0.95

a 0.9 0.99 b

0.95

0.9

The solution can be obtained from a partition of the graph into three columns. Let L denote the event that there is a path of functional devices only through the three units on the left. From the independence and based upon the previous example, $P L( ) = -. 1\ 01^{3}$

Similarly, let M denote the event that there is a path of functional devices only through the two units in the middle. Then, $P M( ) = -. 1\ 0\ 05^{2}$

The probability that there is a path of functional devices only through the one unit on the right is simply the probability that the device functions, namely, 0.99. Therefore, with the independence assumption used again, the solution is 1 0 1 1 0 05 0 99 0 987 $^{3\ 2}$ (¯ ·)( ) − . ( ) . = .

FOR SECTION 2-6

Problem available in *WileyPLUS* at instructor's discretion.

Tutoring problem available in *WileyPLUS* at instructor's discretion

2-142. If $P A B, P B,$ (|) =. 0 4 0 8 and ( ) = . PA, ( ) = . 0 5 are the events A and B independent?

2-143. If $P A B, P B,$ (|) =. 0 3 0 8 and ( ) = . PA, ( ) = . 0 3 are the events B and the complement of A independent?

2-144. If $P A, P B,$ ( ) =. 0 2 0 2 and ( ) = . A and B are mutu ally exclusive, are they independent?

2-145. A batch of 500 containers of frozen orange juice con tains 5 that are defective. Two are selected, at random, without replacement, from the batch. Let A and B denote the events that the fi rst and second containers selected are defective, respectively. (a) Are A and B independent events?

(b) If the sampling were done with replacement, would A and B be independent?

2-146. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

	Shock Resistance
	High Low
Scratch High 70 9 Resistance Low 16 5	

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resist ance. Are events A and B independent?

2-147. Samples of emissions from three suppliers are classi fied for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

	Conforms
	Yes No
	1 22 8
Supplier 2 25 5 3 30 10	

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. (a) Are events A and B independent?

(b) Determine $PB A$ (|).

2-148. Redundant array of inexpensive disks (RAID) is a technology that uses multiple hard drives to increase the speed of data transfer and provide instant data backup. Suppose that the probability of any hard drive failing in a day is 0.001 and the drive failures are independent.

(a) A RAID 0 scheme uses two hard drives, each containing a mirror image of the other. What is the probability of data loss? Assume that data loss occurs if both drives fail within the same day.

(b) A RAID 1 scheme splits the data over two hard drives. What is the probability of data loss? Assume that data loss occurs if at least one drive fails within the same day.

2-149. The probability that a lab specimen contains high levels of contamination is 0.10. Five samples are checked, and the samples are independent.

(a) What is the probability that none contain high levels of contamination?

(b) What is the probability that exactly one contains high lev els of contamination?

(c) What is the probability that at least one contains high levels of contamination?

2-150. In a test of a printed circuit board using a random test pattern, an array of 10 bits is equally likely to be 0 or 1. Assume the bits are independent.

(a) What is the probability that all bits are 1s?

(b) What is the probability that all bits are 0s?
(c) What is the probability that exactly 5 bits are 1s and 5 bits are 0s?

2-151. Six tissues are extracted from an ivy plant infested by spider mites. The plant in infested in 20% of its area. Each tissue is chosen from a randomly selected area on the ivy plant.

Section 2-6/Independence

(a) What is the probability that a player defeats all four oppo nents in a game?
(b) What is the probability that a player defeats at least two opponents in a game?
(c) If the game is played three times, what is the probability that the player defeats all four opponents at least once?

2-153. In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Because acids and bases are usually colorless (as are the water and salt produced in the neutralization reaction), pH is measured to monitor the reaction. Suppose that the equivalence point is reached after approximately 100 mL of an NaOH solu tion has been added (enough to react with all the acetic acid pre sent) but that replicates are equally likely to indicate from 95 to 104 mL, measured to the nearest mL. Assume that two techni cians each conduct titrations independently.
(a) What is the probability that both technicians obtain equiva lence at 100 mL?
(b) What is the probability that both technicians obtain equiva lence between 98 and 104 mL (inclusive)?
(c) What is the probability that the average volume at equiva lence from the technicians is 100 mL?

2-154. A credit card contains 16 digits. It also contains the month and year of expiration. Suppose there are 1 million credit card holders with unique card numbers. A hacker ran domly selects a 16-digit credit card number.
(a) What is the probability that it belongs to a user? (b) Suppose a hacker has a 25% chance of correctly guessing the year your card expires and randomly selects 1 of the 12 months. What is the probability that the hacker correctly selects the month and year of expiration?

2-155. Eight cavities in an injection-molding tool produce plastic connectors that fall into a common stream. A sample is chosen every several minutes. Assume that the samples are independent.
(a) What is the probability that five successive samples were all produced in cavity 1 of the mold?
(b) What is the probability that five successive samples were all produced in the same cavity of the mold?
(c) What is the probability that four out of five successive sam ples were produced in cavity 1 of the mold?

2-156. The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the prob ability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?

(a) What is the probability that four successive samples show the signs of infestation?
(b) What is the probability that three out of four successive samples show the signs of infestation?

2-152. A player of a video game is confronted with a series of four opponents and an 80% probability of defeating each opponent. Assume that the results from opponents are independent (and that when the player is defeated by an oppo nent the game ends).

0.9

0.95
0.8
0.95

0.7
0.95

other devices are functional. What is the probability that the circuit operates?

2-157. The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not

Chapter 2/Probability

0.9
0.95
0.9
0.95

0.8
0.9

Let A and B denote the event that the first bar is wide and B denote the event that the second bar is wide. Determine the following:

(a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$

(d) Are A and B independent events?

2-163. An integrated circuit contains 10 million logic gates (each can be a logical AND or OR circuit). Assume the prob ability of a gate failure is p and that the failures are independent.

2-158. Consider the endothermic reactions in Exercise 2-50. Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is above target. Are these events independent?

2-159. Consider the hospital emergency room data in Example 2-8. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Are these events independent?

2-160. Consider the well failure data in Exercise 2-53. Let A denote the event that the geological formation has more than

1000 wells, and let B denote the event that a well failed. Are these events independent?

2-161. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let A denote the event that the design color is red, and let B denote the event that the font size is not the smallest one. Are A and B independent events? Explain why or why not.

2-162. Consider the code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter).

2-164. Table 2-1 provides data on wafers categorized by location and contam ination levels. Let A denote the event that contamination is *low*, and let B denote the event that the location is *center.* Are A and B independent? Why or why not?

2-165. Table 2-1 provides data on wafers categorized by loca tion and contamination levels. More generally, let the number of wafers with *low* contamination from the *center* and *edge* locations be denoted as n_{lc} and n_{le}, respectively. Similarly, let n_{hc} and n_{he} denote the number of wafers with *high* contamina tion from the *center* and *edge* locations, respectively. Suppose that $n_{lc} = 10n_{hc}$ and $n_{le} = 10n_{he}$. That is, there are 10 times as many *low* con tamination wafers as *high* ones from each loca tion. Let A denote the event that contamination is low, and let B denote the event that the location is *center.* Are A and B inde pendent? Does your conclusion change if the multiplier of 10 (between *low* and *high* contamination wafers) is changed from 10 to another positive integer?

2-7 Bayes' Theorem

The integrated circuit fails to function if any gate fails. Deter mine the value for p so that the probability that the integrated circuit functions is 0.95.

The examples in this chapter indicate that information is often presented in terms of conditional probabilities. These conditional probabilities commonly provide the probability of an event (such as failure) given a condition (such as high or low contamination). But after a random experiment generates an outcome, we are naturally interested in the probability that a condition was present (high contamination) given an outcome (a semiconductor failure). Thomas Bayes addressed this essential question in the 1700s and developed the fundamental result known as **Bayes' theorem**. Do not let the simplicity of the mathematics conceal the importance. There is extensive interest in such probabilities in modern statistical analysis.

From the definition of conditional probability,

$$P(A \cap B) = P(A|B)P(B) = P(B \cap A) = P(B|A)P(A)$$

Now, considering the second and last terms in the preceding expression, we can write

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{for } P(B) > 0 \qquad (2\text{-}15)$$

This is a useful result that enables us to solve for $P(A|B)$ in terms of $P(B|A)$.

Reconsider Example 2-27. The conditional probability that a high level of contamination was present when a failure occurred is to be determined. The information from Example 2-27 is summarized here.

Probability of Failure	Level of Contamination	Probability of Level
0.1	High	0.2
0.005	Not high	0.8

The probability of $P(H \mid F)$ is determined from

$$P(H \mid F) = \frac{P(F \mid H)P(H)}{P(F)} = \frac{0.10(0.20)}{0.024} = 0.83$$

The value of $P(F)$ in the denominator of our solution was found from $P(F) = P(F \mid H)P(H) + P(F \mid H')P(H')$.

In general, if $P(B)$ in the denominator of Equation 2-15 is written using the total probability rule in Equation 2-12, we obtain the following general result, which is known as **Bayes' theorem**.

Bayes' Theorem

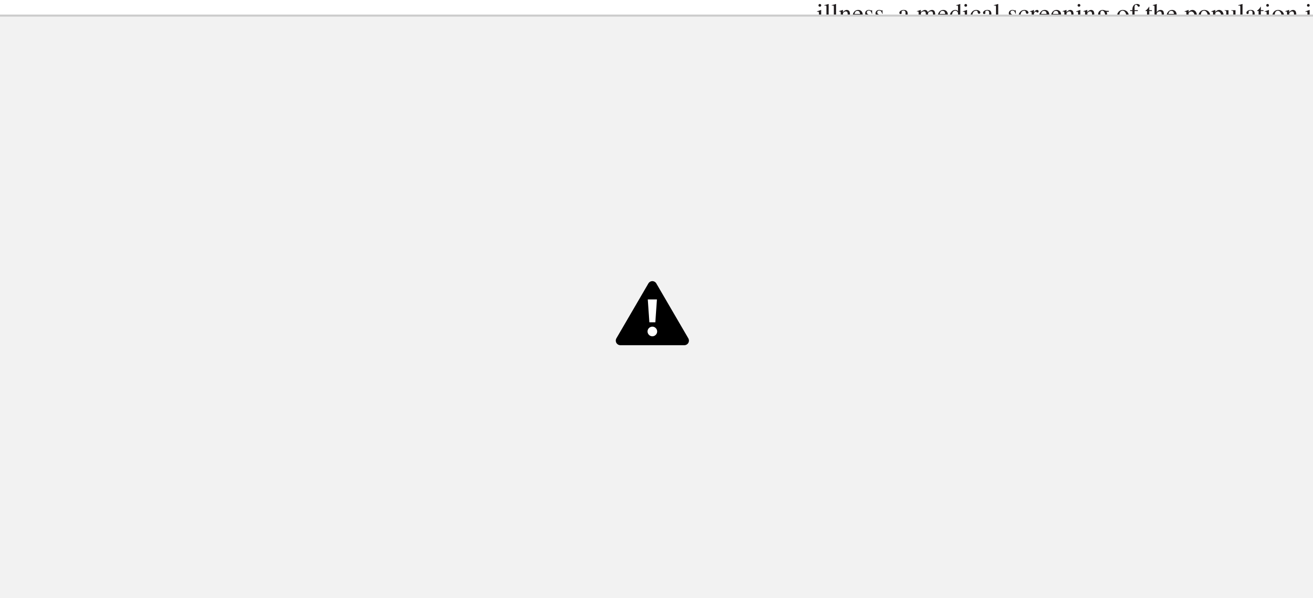
If $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive events and B is any event,

$$P(E_1 \mid B) = \frac{P(B \mid E_1)P(E_1)}{P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k)} \qquad \text{(2-16)}$$

for $P(B) > 0$

in the sum in the denominator.

Medical Diagnostic Because a new medical procedure has been shown to be effective in the early detection of an illness, a medical screening of the population is proposed.



the test correctly identifies someone with the illness as positive is 0.99, and the probability that the test correctly identifies someone without the illness as negative is 0.95. The incidence of the illness in the general population is 0.0001. You take the test, and the result is positive. What is the probability that you have the illness?

Let D denote the event that you have the illness, and let S denote the event that the test signals positive. The probability requested can be denoted as $P(D \mid S)$. The probability that the test correctly signals someone without the illness as negative is 0.95. Consequently, the probability of a positive test without the illness is

PS D ( ) | ′ = . 0 05

From Bayes' theorem, PD S PS DPD PS D D PS D PD ( ) = ( ) ( ) /[( ) ( ) (′) (′)]α

|| | |

P +

= . 0 99 0 0001 [. .

. /

( ) 0001 ( ) (1−)] .+ .

= =. 1 506 0 002 /

0 99 0 0001 0 05 0

Practical Interpretation: The probability of your having the illness given a positive result from the test is only 0.002. Surpris ingly, even though the test is effective, in the sense that *PS D* ( ) | is high and *PS D* ( ) | ′ is low, because the incidence of the illness in the general population is low, the chances are quite small that you actually have the disease even if the test is positive.

Web sites of high-technology manufac turers to allow customers to quickly diagnose problems with products. An

A printer

manufacturer obtained the following probabilities from a database of test results. Printer failures are asso ciated with three types of problems: hardware, software, and other (such as connectors) with probabilities of 0.1, 0.6, and 0.3, respectively. The probability of a printer failure given a hardware problem is 0.9, given a software problem is 0.2, and given any other type of problem is 0.5. If a customer enters the manufacturer's Web site to diagnose a printer failure, what is the most likely cause of the problem?

and chlorinated compounds—instead of having to use a single test for each pollutant. The makers of the test claim that it can detect high levels of organic pollutants with 99.7% accuracy, volatile solvents with 99.95% accuracy, and chlorinated com pounds with 89.7% accuracy. If a pollutant is not present, the test does not signal. Samples are prepared for the calibration of the test and 60% of them are contaminated with organic pollut ants, 27% with volatile solvents, and 13% with traces of chlo rinated compounds. A test sample is selected randomly. (a) What is the probability that the test will signal? (b) If the test signals, what is the probability that chlorinated compounds are present?

2-174. Consider the endothermic reactions in Exercise 2-50. Use Bayes' theorem to calculate the probability that a reac tion's final temperature is 271 K or less given that the heat absorbed is above target.

2-175. Consider the hospital emergency room data in Example 2-8. Use Bayes' theorem to calculate the probability that a person visits hospital 4 given they are LWBS. 2-176. Consider the well failure data in Exercise 2-53. Use Bayes' theorem to calculate the probability that a randomly selected well is in the gneiss group given that the well has failed.

2-177. Two Web colors are used for a site advertisement. If a site visitor arrives from an affiliate, the probabilities of the blue or green colors being used in the advertisement are 0.8 and 0.2, respectively. If the site visitor arrives from a search site, the probabilities of blue and green colors in the advertisement are 0.4 and 0.6, respectively. The proportions of visitors from affiliates and search sites are 0.3

2-8 Random Variables

and 0.7, respectively. What is the probability that a visitor is from a search site given that the blue ad was viewed?

2-178. Suppose that a patient is selected randomly from those described in Exercise 2-98. Let *A* denote the event that the patient is in group 1, and let *B* denote the event that there is no progres sion. Determine the following probabilities:

(a) $P(B)$ (b) $P(B \mid A)$ (c) $P(A \mid B)$ 2-179. An e-mail filter is planned to separate valid e-mails from spam. The word *free* occurs in 60% of the spam mes sages and only 4% of the valid messages. Also, 20% of the messages are spam. Determine the following probabilities: (a) The message contains *free.* (b) The message is spam given that it contains *free.* (c) The message is valid given that it does not contain *free.* 2-180. A recreational equipment supplier finds that among orders that include tents, 40% also include sleeping mats. Only 5% of orders that do not include tents do include sleeping mats. Also, 20% of orders include tents. Determine the following probabilities: (a) The order includes sleeping mats.

(b) The order includes a tent given it includes sleeping mats.

2-181. The probabilities of poor print quality given no printer problem, misaligned paper, high ink viscosity, or printer-head debris are 0, 0.3, 0.4, and 0.6, respectively. The probabilities of no printer problem, misaligned paper, high ink viscosity, or printer-head debris are 0.8, 0.02, 0.08, and 0.1, respectively. (a) Determine the probability of high ink viscosity given poor print quality.
(b) Given poor print quality, what problem is most likely?

Random Variable

Notation

We often summarize the outcome from a random experiment by a simple number. In many of the examples of random experiments that we have considered, the sample space has been a description of possible outcomes. In some cases, descriptions of outcomes are sufficient, but in other cases, it is useful to associate a number with each outcome in the sample space. Because the particular outcome of the experiment is not known in advance, the resulting value of our variable is not known in advance. For this reason, the variable that associates a number with the outcome of a random experiment is referred to as a **random variable**.

A **random variable** is a function that assigns a real number to each outcome in the sample space of a random experiment.

Notation is used to distinguish between a random variable and the real number.

A random variable is denoted by an uppercase letter such as X. After an experiment is conducted, the measured value of the random variable is denoted by a lowercase letter such as $x = 70$ milliamperes.

Sometimes a measurement (such as current in a copper wire or length of a machined part) can assume any value in an interval of real numbers (at least theoretically). Then arbitrary precision in the measurement is possible. Of course, in practice, we might round

off to the nearest tenth or hundredth of a unit. The random variable that represents this measurement is said to be a **continuous random variable**. The range of the random vari able includes all values in an interval of real numbers; that is, the range can be thought of as a continuum.

In other experiments, we might record a count such as the number of transmitted bits that are received in error. Then, the measurement is limited to integers. Or we might record that a proportion such as 0.0042 of the 10,000 transmitted bits were received in error. Then, the measurement is fractional, but it is still limited to discrete points on the real line. Whenever the measurement is limited to discrete points on the real line, the random variable is said to be a **discrete random variable**.

Discrete and Continuous Random Variables

Examples of Random Variables

A **discrete random variable** is a random variable with a finite (or countably infinite) range.

A **continuous random variable** is a random variable with an interval (either finite or infinite) of real numbers for its range.

In some cases, the random variable X is actually discrete but, because the range of possible values is so large, it might be more convenient to analyze X as a continuous random variable. For example, suppose that current measurements are read from a digital instrument that displays the current to the nearest 100th of a milliampere. Because the possible measurements are limited, the random variable is discrete. However, it might be a more convenient, simple approximation to assume that the current measurements are values of a continuous random variable.

Examples of **continuous** random variables:
electrical current, length, pressure, temperature, time, voltage, weight

Examples of **discrete** random variables:
number of scratches on a surface, proportion of defective parts among 1000 tested, number of transmitted bits received in error

FOR SECTION 2-8

Problem available in *WileyPLUS* at instructor's discretion.
Tutoring problem available in *WileyPLUS* at instructor's discretion

2-182. Decide whether a discrete or continuous random variable is the best model for each of the following variables: (a) The time until a projectile returns to earth.
(b) The number of times a transistor in a computer memory changes state in one operation.
(c) The volume of gasoline that is lost to evaporation during the filling of a gas tank.
(d) The outside diameter of a machined shaft.

2-183. Decide whether a discrete or continuous random variable is the best model for each of the following variables:
(a) The number of cracks exceeding one-half inch in 10 miles of an interstate highway.
(b) The weight of an injection-molded plastic part.
(c) The number of molecules in a sample of gas.
(d) The concentration of output from a reactor.
(e) The current in an electronic circuit.

2-184. Decide whether a discrete or continuous random variable is the best model for each of the following variables:
(a) The time for a computer algorithm to assign an image to a category.
(b) The number of bytes used to store a file in a computer. (c) The ozone concentration in micrograms per cubic meter. (d) The ejection fraction (volumetric fraction of blood pumped from a heart ventricle with each beat).
(e) The fluid flow rate in liters per minute.