

6TH EDITION

APPLIED STATISTICS
AND PROBABILITY
FOR ENGINEERS

Douglas C. Montgomery

George C. Runger

WILEY

Index of Applications in Examples and Exercises

BIOLOGICAL

Amino acid composition of soybean meal Exercise 8-52
 Anaerobic respiration Exercise 2-144
 Blood
 Cholesterol level Exercise 15-10
 Glucose level Exercises 13-25, 14-37
 Hypertension Exercises 4-143, 8-31, 11-8, 11-30, 11-46
 Body mass index (BMI) Exercise 11-35
 Body temperature Exercise 9-59
 Cellular replication Exercises 2-193, 3-100
 Circumference of orange trees Exercise 10-46
 Deceased beetles under autolysis and putrefaction Exercise 2-92
 Diet and weight loss Exercises 10-43, 10-77, 15-35
 Disease in plants Exercise 14-76
 Dugongs (sea cows) length Exercise 11-15
 Fatty acid in margarine Exercises 8-36, 8-66, 8-76, 9-147, 9-113
 Gene expression Exercises 6-65, 13-50, 15-42
 Gene occurrence Exercises 2-195, 3-11
 Gene sequences Exercises 2-25, 2-192, 3-13, 3-147
 Grain quality Exercise 8-21
 Height of plants Exercises 4-170, 4-171
 Height or weight of people Exercises 4-44, 4-66, 5-64, 6-30, 6-37, 6-46, 6-63, 6-73, 9-68
 Insect fragments in chocolate bars Exercises 3-134, 4-101
 IQ for monozygotic twins Exercise 10-45
 Leaf transmutation Exercises 2-88, 3-123
 Leg strength Exercises 8-30, 9-64
 Light-dependent photosynthesis Exercise 2-24
 Nisin recovery Exercises 12-14, 12-32, 12-50, 12-64, 12-84, 14-83
 Pesticides and grape infestation Exercise 10-94
 Potato spoilage Exercise 13-14
 Protein
 in Livestock feed Exercise 14-75
 in Milk Exercises 13-13, 13-25, 13-33
 from Peanut milk Exercise 9-143
 Protopectin content in tomatoes Exercises 13-40, 15-40
 Rat muscle Exercise 6-15
 Rat tumors Exercise 8-50
 Rat weight Exercise 8-57
 Rejuvenated mitochondria Exercises 2-96, 3-88
 Root vole population Exercise 14-16
 Sodium content of cornflakes Exercise 9-61
 Soil Exercises 3-24, 12-1, 12-2, 12-23, 12-24, 12-41, 12-42
 Splitting cell Exercise 4-155
 St John's Wort Example 10-14
 Stork sightings Exercises 4-100, 11-96
 Sugar content Exercises 8-46, 9-83, 9-114
 Synapses in the granule cell layer Exercise 9-145
 Tar content in tobacco Exercise 8-95
 Taste evaluation Exercises 14-13, 14-31, 14-34, 14-50, 14-54
 Tissues from an ivy plant Exercise 2-130

Visual accommodation Exercises 6-11, 6-16, 6-75
 Weight of swine or guinea pigs Exercises 9-142, 13-48
 Wheat grain drying Exercises 13-47, 15-41

CHEMICAL

Acid-base titration Exercises 2-60, 2-132, 3-12, 5-48
 Alloys Examples 6-4, 8-5
 Exercises 10-21, 10-44, 10-59, 13-38, 15-17
 Contamination Exercise 2-128, 4-113
 Cooking oil Exercise 2-79
 Etching Exercises 10-19, 10-65, 10-34
 Infrared focal plane arrays Exercise 9-146
 Melting point of a binder Exercise 9-42
 Metallic material transition Examples 8-1, 8-2
 Moisture content in raw material Exercise 3-6
 Mole fraction solubility Exercises 12-75, 12-91
 Mole ratio of sebacic acid Exercise 11-91
 Pitch carbon analysis Exercises 12-10, 12-36, 12-50, 12-60, 12-68
 Plasma etching Examples 14-5, 14-8
 Exercise 7-32
 Polymers Exercises 7-15, 10-8, 13-12, 13-24
 Propellant
 Bond shear strength Examples 15-1, 15-2, 15-4
 Exercises 11-11, 11-31, 11-49, 15-32
 Burning rate Examples 9-1, 9-2, 9-3, 9-4, 9-5
 Exercise 10-6
 Purity Exercise 15-42
 Thermal barrier coatings Exercise 10-75

CHEMICAL ENGINEERING

Aluminum smelting Exercise 10-92
 Automobile basecoat Exercises 14-56, 14-68
 Blow molding Exercise 16-59
 Catalyst usage Exercise 10-17
 Concentration Examples 16-2, 16-6
 Exercises 5-46, 6-68, 6-84, 10-9, 10-54, 15-64
 Conversion Exercise 12-3
 Cooling system in a nuclear submarine Exercise 9-130
 Copper content of a plating bath Exercises 15-8, 15-34, 15-58
 Dispensed syrup in soda machine Exercises 8-29, 8-63, 8-75
 Dry ash value of paper pulp Exercise 14-57
 Fill volume and capability Examples 5-35, 8-6, 9-8, 9-9
 Exercises 2-180, 3-146, 3-151, 4-62, 4-63, 5-62, 9-100, 10-4, 10-85, 10-90, 14-43, 15-38
 Filtration rate Exercise 14-44
 Fish preparation Exercise 13-46
 Flow metering devices Examples 15-3, 15-5
 Exercises 9-126, 9-127
 Foam expanding agents Exercises 10-16, 10-56, 10-88
 Green liquor Exercise 12-100
 Hardwood concentration Example 13-2
 Exercise 14-11

Impurity level in chemical product Exercises 15-3, 15-15
Injection molding Example 14-9
Exercises 2-15, 2-137, 10-70

Laboratory analysis of chemical process samples Exercise 2-43
Maximum heat of a hot tub Exercise 10-33
Na₂S concentration Exercises 11-7, 11-29, 11-41, 11-62
NbOCl₃ Exercise 6-36
Oxygen purity Examples 11-1, 11-2, 11-3, 11-4, 11-5, 11-6, 11-7

pH
and Catalyst concentration Exercise 14-61
of Plating bath Exercises 15-1, 15-13
of a Solution Exercise 6-17
of a Water sample Exercise 2-11
Product color Exercise 14-45
Product solution strength
in recirculation unit Exercise 14-38
Pulp brightness Exercise 13-31
Reaction Time Example 4-5
Exercises 2-13, 2-33, 4-56
Redox reaction experiments Exercise 2-65
Shampoo foam height Exercises 8-91, 9-15, 9-16, 9-17, 9-18, 9-19, 9-128
Stack loss of ammonia Exercises 12-16, 12-34, 12-52, 12-66, 12-85

Temperature
Firing Exercise 13-15
Furnace Exercises 6-55, 6-109
of Hall cell solution Exercise 11-92
Vapor deposition Exercises 13-28, 13-32
Vapor phase oxidation of naphthalene Exercise 6-54
Viscosity Exercises 6-66, 6-88, 6-90, 6-96, 12-73, 12-103, 14-64, 15-20, 15-36, 15-86

Water temperature from power plant cooling tower Exercise 9-40
Water vapor pressure Exercise 11-78
Wine Examples 12-14, 12-15
Exercises 6-35, 6-51

CIVIL ENGINEERING

Cement and Concrete
Hydration Example 10-8
Mixture heat Exercises 9-10, 9-11, 9-12, 9-13, 9-14
Mortar briquettes Exercise 15-79
Strength Exercises 4-57, 15-24
Tensile strength Exercise 15-25
Compressive strength Exercises 13-3, 13-9, 13-19, 14-14, 14-24, 14-48, 7-7, 7-8, 8-13, 8-18, 8-37, 8-69, 8-80, 8-87, 8-90, 15-5

Intrinsic permeability Exercises 11-1, 11-23, 11-39, 11-52
Exercise 3-138, 4-102

Highway pavement cracks Exercise 3-138, 4-102
Pavement deflection Exercises 11-2, 11-16, 11-24, 11-40

Retained strength of asphalt Exercises 13-11, 13-23
Speed limits Exercises 8-59, 10-60
Traffic Exercises 3-87, 3-149, 3-153, 9-190

Wearing seat belts Exercises 10-82, 10-83

COMMUNICATIONS, COMPUTERS, AND NETWORKS

Cell phone signal bars Examples 5-1, 5-3
Cellular neural network speed Exercise 8-39
Code for a wireless garage door Exercise 2-34
Computer clock cycles Exercise 3-8
Computer networks Example 4-21
Exercises 2-10, 2-64, 2-164, 3-148, 3-175, 4-65, 4-94

Corporate Web site errors Exercise 4-84
Digital channel Examples 2-3, 3-4, 3-6, 3-9, 3-12, 3-16, 3-24, 4-15, 5-7, 5-9, 5-10

Electronic messages Exercises 3-158, 4-98, 4-115
Email routes Exercise 2-184
Encryption-decryption system Exercise 2-181
Errors in a communications channel Examples 3-22, 4-17, 4-20
Exercises 2-2, 2-4, 2-46, 3-40, 4-116, 5-5, 5-12, 6-94, 9-135
Exercises 2-81, 2-97, 2-194, 3-91, 3-108
Exercise 10-40

Passwords Exercise 8-82
Programming design languages Exercise 13-49
Response time in computer operation system Exercise 2-45
Software development cost Examples 3-1, 3-14
Telecommunication prefixes Exercises 2-17, 3-2, 3-85, 3-105, 3-132, 3-155, 4-95, 4-105, 4-111, 4-117, 4-160, 5-78, 9-98, 15-9

Transaction processing performance and OLTP benchmark Exercises 2-68, 2-175, 5-10, 5-34, 10-7
Exercise 3-75
Examples 3-25, 5-12, 5-13
Exercises 2-32, 2-191, 3-159, 4-87, 4-140, 5-6

ELECTRONICS

Automobile engine controller Examples 9-10, 9-11
Bipolar transistor current Exercise 14-7
Calculator circuit response Exercises 13-6, 13-18
Circuits Examples 2-35, 7-3
Exercises 2-135, 2-136, 2-170, 2-177, 2-190

Conductivity Exercise 12-105
Current Examples 4-1, 4-5, 4-8, 4-9, 4-12, 16-3
Exercises 10-31, 15-30

Drain and leakage current Exercises 13-41, 11-85
Electromagnetic energy absorption Exercise 10-26
Error recovery procedures Exercises 2-18, 2-166
Inverter transient point Exercises 12-98, 12-99, 12-102
Magnetic tape Exercises 2-189, 3-125
Nickel charge Exercises 2-61, 3-48
Parallel circuits Example 2-34
Power consumption Exercises 6-89, 11-79, 12-6, 12-26, 12-44, 12-58, 12-80
Example 9-13
Exercises 2-3, 9-20, 9-21, 9-22, 9-23, 9-24, 9-28

Printed circuit cards	Example 2-10	Temperature in Phoenix, AZ	Exercise 8-49
Redundant disk array	Exercises 2-42, 3-122	Temperature of sewage discharge	Exercises 6-92, 6-97
Resistors	Exercise 2-127	Voters and air pollution	Exercises 9-27, 9-94
	Example 7-1	Waste water treatment tank	Exercise 2-37
Solder connections	Exercise 6-86	Water demand and quality	Exercises 4-68, 9-137
Strands of copper wire	Exercises 3-1, 15-43, 15-45	Watershed yield	Exercise 11-70
Surface charge	Exercise 2-77		
Surface mount technology (SMT)	Exercise 14-15		
Transistor life	Example 16-5	MATERIALS	
Voltage measurement errors	Exercise 7-51	Baked density of carbon anodes	Exercise 14-4
	Exercise 4-48N	Ceramic substrate	Example 16-4
ENERGY		Coating temperature	Exercises 10-24, 10-60
Consumption in Asia	Exercises 6-29, 6-45, 6-59	Coating weight and surface roughness	Exercise 2-90
Enrichment percentage of reactor fuel rods	Exercises 8-41, 8-71, 8-88	Compressive strength	Exercises 7-56, 11-60
Fuel octane ratings	Exercises 6-22, 6-26, 6-38, 6-42, 6-58, 6-78, 10-7	Flow rate on silicon wafers	Exercises 13-2, 13-16, 15-28
Gasoline cost by month	Exercise 15-98	Insulation ability	Exercise 14-5
Gasoline mileage	Exercises 10-89, 11-6, 11-17, 11-28, 11-44, 11-56, 12-27, 12-55, 12-57, 12-77, 12-89, 15-37	Insulation fluid breakdown time	Exercises 6-8, 6-74
Heating rate index	Exercise 14-46	Izod impact test	Exercises 8-28, 8-62, 8-74, 9-66, 9-80
Petroleum imports	Exercise 6-72	Luminescent ink	Exercise 5-28
Released from cells	Exercise 2-168	Paint drying time	Examples 10-1, 10-2, 10-3 Exercises 14-2, 14-19, 15-8, 15-16
Renewable energy consumption	Exercise 15-78	Particle size	Exercises 4-33, 16-17
Steam usage	Exercises 11-5, 11-27, 11-43, 11-55	Photoresist thickness	Exercise 5-63
Wind power	Exercises 4-132, 11-9	Plastic breaking strength	Exercises 10-5, 10-20, 10-55
ENVIRONMENTAL		Polycarbonate plastic	Example 2-8 Exercises 2-66, 2-76
Arsenic	Example 10-6	Rockwell hardness	Exercises 10-91, 9-115, 15-17
Asbestos	Exercises 12-12, 12-30, 12-48, 12-62, 12-76, 12-88, 13-39	Temperature of concrete	Exercise 9-58
Biochemical oxygen demand (BOD)	Exercises 4-85, 4-169	Tensile strength of	
Calcium concentration in lake water	Exercises 11-13, 11-33, 11-51	Aluminum	Example 10-4
Carbon dioxide in the atmosphere	Exercise 8-9	Fiber	Exercises 7-3, 7-4, 13-3, 13-17
Chloride in surface streams	Exercise 3-58	Steel	Example 10-9 Exercise 9-44
Cloud seeding	Exercises 11-10, 11-32, 11-48, 11-59	Paper	Example 13-1 Exercises 4-154, 11-86
Earthquakes	Exercise 9-60	Titanium content	Exercises 8-47, 9-79, 15-2, 15-12
Emissions and fluoride emissions	Exercises 6-63, 9-102, 11-15, 15-46	Tube brightness in TV sets	Exercises 7-12, 8-35, 8-67, 8-79, 9-148, 9-67, 14-1
Global temperature	Exercises 2-28, 15-34	MECHANICAL	
Hydrophobic organic substances	Exercises 6-83, 11-74	Aircraft manufacturing	Examples 6-6, 12-12, 14-1, 15-6, 16-1
Mercury contamination	Exercise 10-93	Artillery shells	Exercises 6-8, 8-97, 10-42, 15-31, 15-13, 15-74
Ocean wave height	Example 8-4	Beam delamination	Exercise 9-106
Organic pollution	Exercise 4-181	Bearings	Exercises 8-32, 8-64
Oxygen concentration	Example 3-18	Diameter	Examples 8-7, 8-8 Exercise 9-95
Ozone levels	Exercises 8-94, 9-63, 9-140	Wear	Exercises 4-181, 9-42, 15-6, 15-14
Radon release	Exercises 2-9, 11-90		Example 4-25 Exercises 5-22, 4-127, 12-19, 12-39, 12-45, 12-67
Rainfall in Australia	Exercises 13-8, 13-20		
Suspended solids in lake water	Exercises 8-33, 8-65, 8-77		
	Exercises 6-32, 6-48, 6-60, 6-80, 9-70		



Applied Statistics and Probability for Engineers

Sixth Edition

Douglas C. Montgomery

Arizona State University

George C. Runger

Arizona State University

WILEY



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.

Copyright © 2014, 2011, 2008, 2005, John Wiley & Sons, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA 01923, website www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201)748-6011, fax (201)748-6008, website <http://www.wiley.com/go/permissions>.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at www.wiley.com/go/returnlabel. Outside of the United States, please contact your local representative.

Library of Congress Cataloging-in-Publication Data

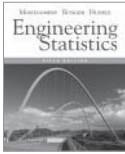
ISBN-13 9781118539712
ISBN (BRV)-9781118645062

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 either 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*

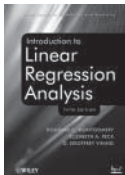
For a first course in statistical quality control. A comprehensive treatment of statistical methodology for quality control and improvement. Some aspects of quality management are also included such as the six-sigma approach.



Design and Analysis of Experiments, Eighth Edition

by *Douglas C. Montgomery*

An introduction to design and analysis of experiments, 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 *Montgomery, 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.

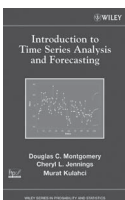


Generalized Linear Models: with Applications in Engineering and the Sciences, Second Edition

by *Myers, Montgomery and Vining*

Website: www.wiley.com/college/myers

An introductory text or reference on generalized linear 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 data and generate forecasts useful to the decision maker. Minitab and SAS are used to illustrate how the methods are implemented in practice. For Advanced undergrad/first-year graduate, with a prerequisite of basic statistical methods. Portions of the book require calculus and matrix algebra.

WileyPLUS

WileyPLUS is a research-based online environment for effective teaching and learning.

WileyPLUS builds students' confidence because it takes the guesswork out of studying by providing students with a clear roadmap:

- what to do
- how to do it
- if they did it right

It offers interactive resources along with a complete digital textbook that help students learn more. With *WileyPLUS*, students take more initiative so you'll have greater impact on their achievement in the classroom and beyond.



Now available for



Blackboard

For more information, visit www.wileyplus.com

WileyPLUS

ALL THE HELP, RESOURCES, AND PERSONAL SUPPORT YOU AND YOUR STUDENTS NEED!

www.wileyplus.com/resources

1st DAY OF
CLASS
... AND BEYOND!

2-Minute Tutorials and all of the resources you and your students need to get started

WileyPLUS

Student
Partner
Program

Student support from an experienced student user

Wiley Faculty Network



Collaborate with your colleagues, find a mentor, attend virtual and live events, and view resources
www.WhereFacultyConnect.com

WileyPLUS

Quick
Start

Pre-loaded, ready-to-use assignments and presentations created by subject matter experts



Technical Support 24/7
FAQs, online chat,
and phone support
www.wileyplus.com/support



© Courtney Keating/
iStockphoto

Your *WileyPLUS* Account Manager,
providing personal training
and support

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 individuals 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 reorganizing 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 methodology 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 summary 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 distributions, the central limit theorem, and point estimation of parameters. This chapter also introduces some of the important properties of estimators, the method of maximum likelihood, 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 tolerance 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 appropriate 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, mathematical 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 adequacy 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 fractional factorials are emphasized. Chapter 15 introduces statistical quality control, emphasizing 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.

- Because computer intensive methods are so important in the modern use of statistics, we have added material on the bootstrap and its use in constructing confidence intervals.
- We have increased the emphasis on the use of P -value in hypothesis testing. Many sections of several chapters were rewritten to reflect this.
- Many sections of the book have been edited and rewritten to improve the explanations and try to make the concepts easier to understand.
- The introductory chapter on hypothesis testing now includes coverage of equivalence testing, a technique widely used in the biopharmaceutical industry, but which has widespread applications in other areas.
- Combining P -values when performing multiple tests is included.
- Decision theory is briefly introduced in Chapter 15.
- We have added brief comments at the end of examples to emphasize the practical interpretations of the results.
- Many new examples and homework exercises have been added.

FEATURED IN THIS BOOK

Definitions, Key Concepts, and Equations

Throughout the text, definitions and key concepts and equations are highlighted by a box to emphasize their importance.

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.

Learning Objectives

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

1. Determine probabilities from probability density functions
2. Determine probabilities from cumulative distribution functions and cumulative distribution functions from probability density functions, and the reverse
3. Calculate means and variances for continuous random variables
4. Understand the assumptions for some common continuous probability distributions
5. Select an appropriate continuous probability distribution to calculate probabilities in specific applications
6. Calculate probabilities, determine means and variances for some common continuous probability distributions
7. Standardize normal random variables
8. Use the table for the cumulative distribution function of a standard normal distribution to calculate probabilities
9. Approximate probabilities for some binomial and Poisson distributions

4-3 Cumulative Distribution Functions

An alternative method to describe the distribution of a discrete random variable can also be used for continuous random variables.

Cumulative Distribution Function

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du \quad (4-3)$$

for $-\infty < x < \infty$.

The cumulative distribution function is defined for all real numbers. The following example illustrates the definition.

Example 4-3 Electric Current For the copper current measurement in Example 4-1, the cumulative distribution function of the random variable X consists of three expressions. If $x < 4.9$, $f(x) = 0$.

Therefore,

$$F(x) = 0, \quad \text{for } x < 4.9$$

and

$$F(x) = \int_{4.9}^x f(u) du = 5x - 24.5, \quad \text{for } 4.9 \leq x < 5.1$$

Finally,

$$F(x) = \int_{4.9}^x f(u) du = 1, \quad \text{for } 5.1 \leq x$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 4.9 \\ 5x - 24.5 & 4.9 \leq x < 5.1 \\ 1 & 5.1 \leq x \end{cases}$$

The plot of $F(x)$ is shown in Fig. 4-6.

Notice that in the definition of $F(x)$, any $<$ can be changed to \leq and vice versa. That is, in Example 4-3 $F(x)$ can be defined as either $5x - 24.5$ or 0 at the end-point $x = 4.9$, and $F(x)$ can be defined as either $5x - 24.5$ or 1 at the end-point $x = 5.1$. In other words, $F(x)$ is a continuous function. For a discrete random variable, $F(x)$ is not a continuous function. Sometimes a continuous random variable is defined as one that has a continuous cumulative distribution function.

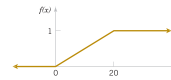


FIGURE 4-6 Cumulative distribution function for Example 4-3.



FIGURE 4-7 Cumulative distribution function for Example 4-4.

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.

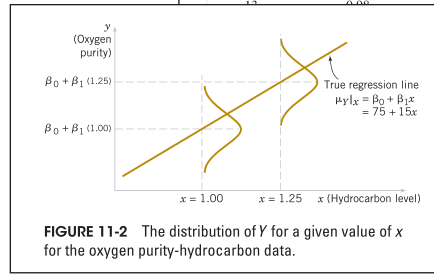
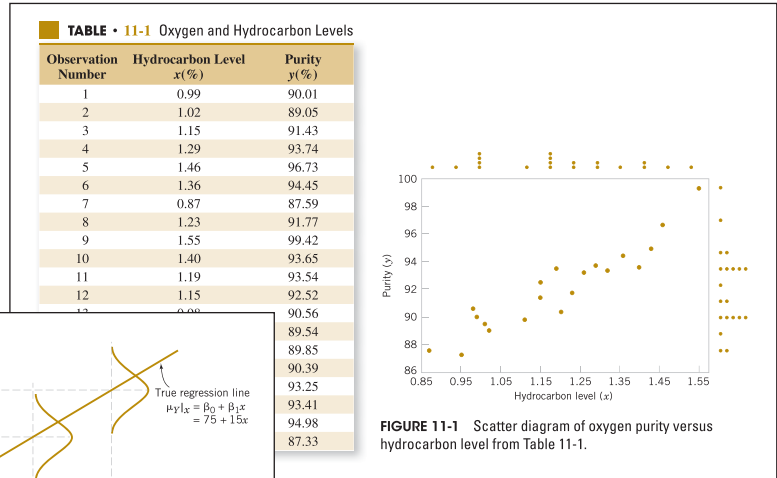
9-1.6 GENERAL PROCEDURE FOR HYPOTHESIS TESTS

This chapter develops hypothesis-testing procedures for many practical problems. Use of the following sequence of steps in applying hypothesis-testing methodology is recommended.

1. **Parameter of interest:** From the problem context, identify the parameter of interest.
2. **Null hypothesis, H_0 :** State the null hypothesis, H_0 .
3. **Alternative hypothesis, H_1 :** Specify an appropriate alternative hypothesis, H_1 .
4. **Test statistic:** Determine an appropriate test statistic.
5. **Reject H_0 if:** State the rejection criteria for the null hypothesis.
6. **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
7. **Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

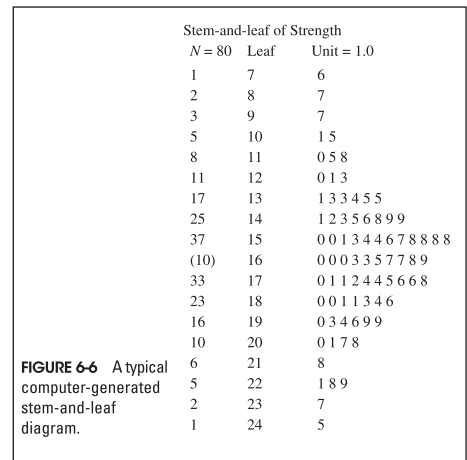
Figures

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.

Example 10-1 Paint Drying Time A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested: formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce the drying time. From experience, it is known that the standard deviation of drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2; the 20 specimens are painted in random order. The two sample average drying times are $\bar{x}_1 = 121$ minutes and $\bar{x}_2 = 112$ minutes, respectively. What conclusions can the product developer draw about the effectiveness of the new ingredient, using $\alpha = 0.05$?

We apply the seven-step procedure to this problem as follows:

- Parameter of interest:** The quantity of interest is the difference in mean drying times, $\mu_1 - \mu_2$, and $\Delta_0 = 0$.
- Non hypothesis:** $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 = \mu_2$.
- Alternative hypothesis:** $H_1: \mu_1 > \mu_2$. We want to reject H_0 if the new ingredient reduces mean drying time.
- Test statistic:** The test statistic is

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{where } \sigma_1^2 = \sigma_2^2 = (8)^2 = 64 \text{ and } n_1 = n_2 = 10.$$
- Reject H_0 if:** Reject $H_0: \mu_1 = \mu_2$ if the P -value is less than 0.05.
- Computations:** Because $\bar{x}_1 = 121$ minutes and $\bar{x}_2 = 112$ minutes, the test statistic is

$$z_0 = \frac{121 - 112}{\sqrt{\frac{(8)^2}{10} + \frac{(8)^2}{10}}} = 2.52$$
- Conclusion:** Because $z_0 = 2.52$, the P -value is $P = 1 - \Phi(2.52) = 0.0059$, so we reject H_0 at the $\alpha = 0.05$ level.

Practical Interpretation We conclude that adding the new ingredient to the paint significantly reduces the drying time. This is a strong conclusion.

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.

EXERCISES FOR SECTION 5-5

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

5-79. + Suppose that X is a random variable with probability distribution $f_X(x) = 1/4, \quad x = 1, 2, 3, 4$

5-85. + Suppose that X has the probability distribution $f_X(x) = 1, \quad 1 \leq x \leq 2$. Determine the probability distribution of the random variable $Y = e^X$.

Supplemental Exercises

5-101. + Show that the following function satisfies the properties of a joint probability mass function:

X	Y	$f(x, y)$
0	0	1/4
0	1	1/8
1	0	1/8
1	1	1/4

(e) Conditional mean of the number of people who suffer severe side effects given that 19 suffer minor side effects.

5-103. The backoff torque required to remove bolts in a steel plate is rated as high, moderate, or low. Historically, the probability of a high, moderate, or low rating is 0.6, 0.3, or 0.1, respectively. Suppose that 20 bolts are evaluated and that the torque ratings are independent.

Mind-expanding exercises

5-126. Show that if X_1, X_2, \dots, X_p are independent, continuous random variables, $P(X_1 \in A_1, X_2 \in A_2, \dots, X_p \in A_p) = P(X_1 \in A_1)P(X_2 \in A_2) \dots P(X_p \in A_p)$ for any regions A_1, A_2, \dots, A_p in the range of X_1, X_2, \dots, X_p respectively.

5-127. Show that if X_1, X_2, \dots, X_p are independent random variables and $Y = c_1X_1 + c_2X_2 + \dots + c_pX_p$, $V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \dots + c_p^2V(X_p)$.

You may assume that the random variables are continuous.

5-128. Suppose that the joint probability function of the continuous random variables X and Y is constant on the rectangle $0 < x < a$, $0 < y < b$. Show that X and Y are independent.

5-129. Suppose that the range of the continuous variables X and Y is $0 < x < a$ and $0 < y < b$. Also suppose that the joint probability density function $f_{X,Y}(x, y) = g(x)h(y)$, where $g(x)$ is a function only of x , and $h(y)$ is a function only of y . Show that X and Y are independent.

5-130. This exercise extends the hypergeometric distribution to multiple variables. Consider a population with N items of k different types. Assume that there are N_1 items of type 1, N_2 items of type 2, ..., N_k items of type k so that $N_1 + N_2 + \dots + N_k = N$. Suppose that a random sample of size n is selected, without replacement, from the population. Let X_1, X_2, \dots, X_k denote the number of items of each type in the sample so that $X_1 + X_2 + \dots + X_k = n$. Show that for feasible values of $n, x_1, x_2, \dots, x_k, N_1, N_2, \dots, N_k$, the probability is

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{\binom{N_1}{x_1} \binom{N_2}{x_2} \dots \binom{N_k}{x_k}}{\binom{N}{n}}$$

5-131. Use the properties of moment generating functions to show that a sum of p independent normal random variables with means μ_i and variances σ_i^2 for $i=1, 2, \dots, p$ has a normal distribution.

5-132. Show that by expanding e^{tX} in a power series and taking expectations term by term we may write the moment-generating function as

$$M_X(t) = E(e^{tX}) = 1 + \mu_1' t + \frac{\mu_2' t^2}{2!} + \dots + \frac{\mu_r' t^r}{r!} + \dots$$

Thus, the coefficient of $t^r / r!$ in this expansion is μ_r' , the r th origin moment.

Write the power series expansion for $M_X(t)$ for a gamma random variable and determine μ_1' and μ_2' using this approach.

Important Terms and Concepts

Bivariate distribution	Contour plots	Joint probability density function	Marginal probability distribution
Bivariate normal distribution	Correlation	Joint probability distribution function	Moment-generating functions
Conditional mean	Covariance	Joint probability mass function	Multinomial distribution
Conditional probability density function	Error propagation	Linear functions of random variables	Reproductive property of the normal distribution
Conditional probability mass function	General functions of random variables		
Conditional variance	Independence		

STUDENT RESOURCES

- **Data Sets** Data sets for all examples and exercises in the text. Visit the student section of the book Web site at www.wiley.com/college/montgomery to access these materials.
- **Student Solutions Manual** Detailed solutions for selected problems in the book. The *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*:

- Students achieve concept mastery in a rich, structured environment that's available 24/7.
- Instructors personalize and manage their course more effectively with assessment, assignments, 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:

- Students receive timely access to resources that address their demonstrated needs, and get immediate feedback and remediation when needed.
- Integrated, multi-media resources—including audio and visual exhibits, demonstration problems, and much more—provide multiple study-paths to fit each student's learning preferences 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:

- You can customize your classroom presentation with a wealth of resources and functionality from PowerPoint slides to a database of rich visuals. You can even add your own materials to your *WileyPLUS* course.
- With *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, making assignments, scoring student work, keeping grades, and more.

COURSE SYLLABUS SUGGESTIONS

This is a very flexible textbook because instructors' ideas about what should be in a first course on statistics for engineers vary widely, as do the abilities of different groups of students. 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 concepts. 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, Statgraphics, 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 ultimately 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 precision 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 Borrer

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

Sarah Streett Yolande Tra

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 *Biometrics* allowed us to use copyrighted material, for which we are grateful.

Douglas C. Montgomery

George C. Runger

Contents

Inside Front cover Index of Applications Examples and Exercises

Chapter 1 The Role of Statistics in Engineering 1

- 1-1 The Engineering Method and Statistical Thinking 2
- 1-2 Collecting Engineering Data 4
 - 1-2.1 Basic Principles 4
 - 1-2.2 Retrospective Study 5
 - 1-2.3 Observational Study 5
 - 1-2.4 Designed Experiments 6
 - 1-2.5 Observing Processes Over Time 8
- 1-3 Mechanistic and Empirical Models 11
- 1-4 Probability and Probability Models 12

Chapter 2 Probability 15

- 2-1 Sample Spaces and Events 16
 - 2-1.1 Random Experiments 16
 - 2-1.2 Sample Spaces 17
 - 2-1.3 Events 20
 - 2-1.4 Counting Techniques 22
- 2-2 Interpretations and Axioms of Probability 30
- 2-3 Addition Rules 35
- 2-4 Conditional Probability 40
- 2-5 Multiplication and Total Probability Rules 45
- 2-6 Independence 49
- 2-7 Bayes' Theorem 54
- 2-8 Random Variables 57

Chapter 3 Discrete Random Variables and Probability Distributions 65

- 3-1 Discrete Random Variables 66
- 3-2 Probability Distributions and Probability Mass Functions 67
- 3-3 Cumulative Distribution Functions 71
- 3-4 Mean and Variance of a Discrete Random Variable 74
- 3-5 Discrete Uniform Distribution 78
- 3-6 Binomial Distribution 80
- 3-7 Geometric and Negative Binomial Distributions 86
 - 3-7.1 Geometric Distribution 86
- 3-8 Hypergeometric Distribution 93
- 3-9 Poisson Distribution 98

Chapter 4 Continuous Random Variables and Probability Distributions 107

- 4-1 Continuous Random Variables 108
- 4-2 Probability Distributions and Probability Density Functions 108
- 4-3 Cumulative Distribution Functions 112
- 4-4 Mean and Variance of a Continuous Random Variable 114
- 4-5 Continuous Uniform Distribution 116
- 4-6 Normal Distribution 119
- 4-7 Normal Approximation to the Binomial and Poisson Distributions 128
- 4-8 Exponential Distribution 133
- 4-9 Erlang and Gamma Distributions 139
- 4-10 Weibull Distribution 143
- 4-11 Lognormal Distribution 145
- 4-12 Beta Distribution 148

Chapter 5 Joint Probability Distributions 155

- 5-1 Two or More Random Variables 156
 - 5-1.1 Joint Probability Distributions 156
 - 5-1.2 Marginal Probability Distributions 159
 - 5-1.3 Conditional Probability Distributions 161
 - 5-1.4 Independence 164
 - 5-1.5 More Than Two Random Variables 167
- 5-2 Covariance and Correlation 174
- 5-3 Common Joint Distributions 179
 - 5-3.1 Multinomial Probability Distribution 179
 - 5-3.2 Bivariate Normal Distribution 181
- 5-4 Linear Functions of Random Variables 184
- 5-5 General Functions of Random Variables 188
- 5-6 Moment-Generating Functions 191

Chapter 6 Descriptive Statistics 199

- 6-1 Numerical Summaries of Data 200
- 6-2 Stem-and-Leaf Diagrams 206
- 6-3 Frequency Distributions and Histograms 213
- 6-4 Box Plots 217
- 6-5 Time Sequence Plots 219
- 6-6 Scatter Diagrams 225
- 6-7 Probability Plots 230

Chapter 7 Point Estimation of Parameters and Sampling Distributions 239

- 7-1 Point Estimation 240

- 7-2 Sampling Distributions and the Central Limit Theorem 241
 - 7-3 General Concepts of Point Estimation 249
 - 7-3.1 Unbiased Estimators 249
 - 7-3.2 Variance of a Point Estimator 251
 - 7-3.3 Standard Error: Reporting a Point Estimate 251
 - 7.3.4 Bootstrap Standard Error 252
 - 7-3.5 Mean Squared Error of an Estimator 254
 - 7-4 Methods of Point Estimation 256
 - 7-4.1 Method of Moments 256
 - 7-4.2 Method of Maximum Likelihood 258
 - 7-4.3 Bayesian Estimation of Parameters 264
- Chapter 8 Statistical Intervals for a Single Sample 271**
-
- 8-1 Confidence Interval on the Mean of a Normal Distribution, Variance Known 273
 - 8-1.1 Development of the Confidence Interval and Its Basic Properties 273
 - 8-1.2 Choice of Sample Size 276
 - 8-1.3 One-Sided Confidence Bounds 277
 - 8-1.4 General Method to Derive a Confidence Interval 277
 - 8-1.5 Large-Sample Confidence Interval for μ 279
 - 8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown 282
 - 8-2.1 t Distribution 283
 - 8-2.2 t Confidence Interval on μ 284
 - 8-3 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution 287
 - 8-4 Large-Sample Confidence Interval for a Population Proportion 291
 - 8-5 Guidelines for Constructing Confidence Intervals 296
 - 8.6 Bootstrap Confidence Interval 296
 - 8-7 Tolerance and Prediction Intervals 297
 - 8-7.1 Prediction Interval for a Future Observation 297
 - 8-7.2 Tolerance Interval for a Normal Distribution 298
- Chapter 9 Tests of Hypotheses for a Single Sample 305**
-
- 9-1 Hypothesis Testing 306
 - 9-1.1 statistical hypotheses 306
 - 9-1.2 Tests of Statistical Hypotheses 308
 - 9-1.3 One-Sided and Two-Sided Hypotheses 313
 - 9-1.4 P -Values in Hypothesis Tests 314
 - 9-1.5 Connection Between Hypothesis Tests and Confidence Intervals 316
 - 9-1.6 General Procedure for Hypothesis Tests 318
 - 9-2 Tests on the Mean of a Normal Distribution, Variance Known 322
 - 9-2.1 Hypothesis Tests on the Mean 322
 - 9-2.2 Type II Error and Choice of Sample Size 325
 - 9-2.3 Large-Sample Test 329
 - 9-3 Tests on the Mean of a Normal Distribution, Variance Unknown 331
 - 9-3.1 Hypothesis Tests on the Mean 331
 - 9-3.2 Type II Error and Choice of Sample Size 336
 - 9-4 Tests on the Variance and Standard Deviation of a Normal Distribution 340
 - 9-4.1 Hypothesis Tests on the Variance 341
 - 9-4.2 Type II Error and Choice of Sample Size 343
 - 9-5 Tests on a Population Proportion 344
 - 9-5.1 Large-Sample Tests on a Proportion 344
 - 9-5.2 Type II Error and Choice of Sample Size 347
 - 9-6 Summary Table of Inference Procedures for a Single Sample 350
 - 9-7 Testing for Goodness of Fit 350
 - 9-8 Contingency Table Tests 354
 - 9-9 Nonparametric Procedures 357
 - 9-9.1 The Sign Test 358
 - 9-9.2 The Wilcoxon Signed-Rank Test 362
 - 9-9.3 Comparison to the t -Test 364
 - 9-10 Equivalence Testing 365
 - 9-11 Combining P -Values 367
- Chapter 10 Statistical Inference for Two Samples 373**
-
- 10-1 Inference on the Difference in Means of Two Normal Distributions, Variances Known 374
 - 10-1.1 Hypothesis Tests on the Difference in Means, Variances Known 376
 - 10-1.2 Type II Error and Choice of Sample Size 377
 - 10-1.3 Confidence Interval on the Difference in Means, Variances Known 379
 - 10-2 Inference on the Difference in Means of two Normal Distributions, Variances Unknown 383

- 10-2.1 Hypotheses Tests on the Difference in Means, Variances Unknown 383
- 10-2.2 Type II Error and Choice of Sample Size 389
- 10-2.3 Confidence Interval on the Difference in Means, Variances Unknown 390
- 10-3 A Nonparametric Test for the Difference in Two Means 396
 - 10-3.1 Description of the Wilcoxon Rank-Sum Test 397
 - 10-3.2 Large-Sample Approximation 398
 - 10-3.3 Comparison to the t -Test 399
- 10-4 Paired t -Test 400
- 10-5 Inference on the Variances of Two Normal Distributions 407
 - 10-5.1 F Distribution 407
 - 10-5.2 Hypothesis Tests on the Ratio of Two Variances 409
 - 10-5.3 Type II Error and Choice of Sample Size 411
 - 10-5.4 Confidence Interval on the Ratio of Two Variances 412
- 10-6 Inference on Two Population Proportions 414
 - 10-6.1 Large-Sample Tests on the Difference in Population Proportions 414
 - 10-6.2 Type II Error and Choice of Sample Size 416
 - 10-6.3 Confidence Interval on the Difference in Population Proportions 417
- 10-7 Summary Table and Road Map for Inference Procedures for Two Samples 420

Chapter 11 Simple Linear Regression and Correlation 427

- 11-1 Empirical Models 428
- 11-2 Simple Linear Regression 431
- 11-3 Properties of the Least Squares Estimators 440
- 11-4 Hypothesis Tests in Simple Linear Regression 441
 - 11-4.1 Use of t -Tests 441
 - 11-4.2 Analysis of Variance Approach to Test Significance of Regression 443
- 11-5 Confidence Intervals 447
 - 11-5.1 Confidence Intervals on the Slope and Intercept 447
 - 11-5.2 Confidence Interval on the Mean Response 448
- 11-6 Prediction of New Observations 449
- 11-7 Adequacy of the Regression Model 452
 - 11-7.1 Residual Analysis 453
 - 11-7.2 Coefficient of Determination (R^2) 454
- 11-8 Correlation 457
- 11-9 Regression on Transformed Variables 463
- 11-10 Logistic Regression 467

Chapter 12 Multiple Linear Regression 477

- 12-1 Multiple Linear Regression Model 478
 - 12-1.1 Introduction 478
 - 12-1.2 Least Squares Estimation of the Parameters 481
 - 12-1.3 Matrix Approach to Multiple Linear Regression 483
 - 12-1.4 Properties of the Least Squares Estimators 488
- 12-2 Hypothesis Tests In Multiple Linear Regression 497
 - 12-2.1 Test for Significance of Regression 497
 - 12-2.2 Tests on Individual Regression Coefficients and Subsets of Coefficients 500
- 12-3 Confidence Intervals In Multiple Linear Regression 506
 - 12-3.1 Confidence Intervals on Individual Regression Coefficients 506
 - 12-3.2 Confidence Interval on the Mean Response 507
- 12-4 Prediction of New Observations 508
- 12-5 Model Adequacy Checking 511
 - 12-5.1 Residual Analysis 511
 - 12-5.2 Influential Observations 514
- 12-6 Aspects of Multiple Regression Modeling 517
 - 12-6.1 Polynomial Regression Models 517
 - 12-6.2 Categorical Regressors and Indicator Variables 519
 - 12-6.3 Selection of Variables and Model Building 522
 - 12-6.4 Multicollinearity 529

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

- 13-1 Designing Engineering Experiments 540
- 13-2 Completely Randomized Single-Factor Experiment 541
 - 13-2.1 Example: Tensile Strength 541
 - 13-2.2 Analysis of Variance 542
 - 13-2.3 Multiple Comparisons Following the ANOVA 549

- 13-2.4 Residual Analysis and Model Checking 551
- 13-2.5 Determining Sample Size 553
- 13-3 The Random-Effects Model 559
 - 13-3.1 Fixed Versus Random Factors 559
 - 13-3.2 ANOVA and Variance Components 560
- 13-4 Randomized Complete Block Design 565
 - 13-4.1 Design and Statistical Analysis 565
 - 13-4.2 Multiple Comparisons 570
 - 13-4.3 Residual Analysis and Model Checking 571

Chapter 14 Design of Experiments with Several Factors 575

- 14-1 Introduction 576
- 14-2 Factorial Experiments 578
- 14-3 Two-Factor Factorial Experiments 582
 - 14-3.1 Statistical Analysis of the Fixed-Effects Model 582
 - 14-3.2 Model Adequacy Checking 587
 - 14-3.3 One Observation per Cell 588
- 14-4 General Factorial Experiments 591
- 14-5 2^k Factorial Designs 594
 - 14-5.1 2^2 Design 594
 - 14-5.2 2^k Design for $k \geq 3$ Factors 600
 - 14-5.3 Single Replicate of the 2^k Design 607
 - 14-5.4 Addition of Center Points to a 2^k Design 611
- 14-6 Blocking and Confounding in the 2^k Design 619
- 14-7 Fractional Replication of the 2^k Design 626
 - 14-7.1 One-Half Fraction of the 2^k Design 626
 - 14-7.2 Smaller Fractions: The 2^{k-p} Fractional Factorial 632
- 14-8 Response Surface Methods and Designs 643

Chapter 15 Statistical Quality Control 663

- 15-1 Quality Improvement and Statistics 664
 - 15-1.1 Statistical Quality Control 665
 - 15-1.2 Statistical Process Control 666
- 15-2 Introduction to Control Charts 666
 - 15-2.1 Basic Principles 666
 - 15-2.2 Design of a Control Chart 670
 - 15-2.3 Rational Subgroups 671
 - 15-2.4 Analysis of Patterns on Control Charts 672
- 15-3 \bar{X} and R or S Control Charts 674
- 15-4 Control Charts for Individual Measurements 684

- 15-5 Process Capability 692
- 15-6 Attribute Control Charts 697
 - 15-6.1 P Chart (Control Chart for Proportions) 697
 - 15-6.2 U Chart (Control Chart for Defects per Unit) 699
- 15-7 Control Chart Performance 704
- 15-8 Time-Weighted Charts 708
 - 15-8.1 Cumulative Sum Control Chart 709
 - 15-8.2 Exponentially Weighted Moving-Average Control Chart 714
- 15-9 Other SPC Problem-Solving Tools 722
- 15-10 Decision Theory 723
 - 15-10.1 Decision Models 723
 - 15-10.2 Decision Criteria 724
- 15-11 Implementing SPC 726

Appendix A. Statistical Tables and Charts 737

- Table I Summary of Common Probability Distributions 738
- Table II Cumulative Binomial Probabilities $P(X \leq x)$ 739
- Table III Cumulative Standard Normal Distribution 742
- Table IV Percentage Points $\chi^2_{\alpha, v}$ of the Chi-Squared Distribution 744
- Table V Percentage Points $t_{\alpha, v}$ of the t Distribution 745
- Table VI Percentage Points f_{α, v_1, v_2} of the F Distribution 746
- Chart VII Operating Characteristic Curves 751
- Table VIII Critical Values for the Sign Test 760
- Table IX Critical Values for the Wilcoxon Signed-Rank Test 760
- Table X Critical Values for the Wilcoxon Rank-Sum Test 761
- Table XI Factors for Constructing Variables Control Charts 762
- Table XII Factors for Tolerance Intervals 762

Appendix B: Bibliography 765

Appendix C: Answers to Selected Exercises 769

Glossary 787

Index 803

Index of applications in examples and exercises, continued 809



© babyblueut / iStockphoto

The Role of Statistics in Engineering

Chapter Outline

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 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 field are concerned about the capacity of regional highway systems. A typical problem related to transportation would involve data regarding this specific system's number of nonwork, home-based trips, the number of persons per household, and the number of vehicles 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 statistical technique called *regression analysis* can be used to construct 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 engineering. 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 engineering 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.

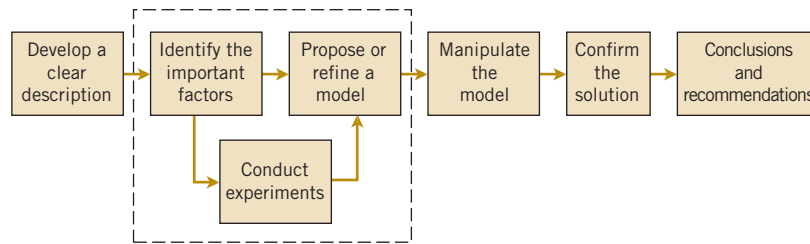


FIGURE 1-1 The engineering method.

The Science of Data

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 science 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.

Variability

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, consider 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 compression, 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 variability 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

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

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 disturbances, 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.



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

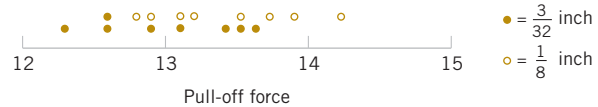


FIGURE 1-3 Dot diagram of pull-off force for two wall thicknesses.

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 prototypes 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 decision? 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?

Population and Samples

Often, physical laws (such as Ohm's law and the ideal gas law) are applied to help design products 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. Sometimes 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**

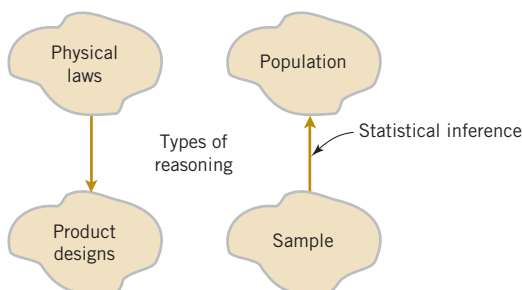


FIGURE 1-4 Statistical inference is one type of reasoning.

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 variables 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 **outliers** (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 controllable 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 discovering 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 scientific or engineering theory is directly or completely applicable, so experimentation and observation 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 theory 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 connector. 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 concentration. 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 levels, “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

TABLE • 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

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

**Interaction can be a
Key Element in
Problem Solving**

An important advantage of factorial experiments is that they allow one to detect an **interaction** 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 temperature 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, condensate temperature, reflux 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 rapidly; 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 five 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 figure 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.

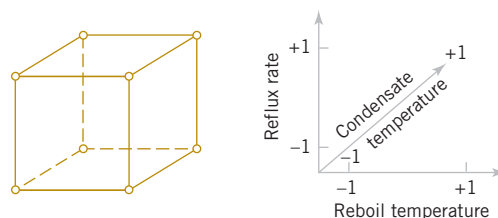
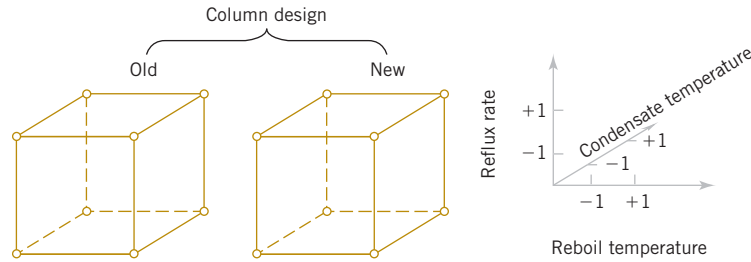


FIGURE 1-5 The factorial design for the distillation column.

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



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 engineering 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.

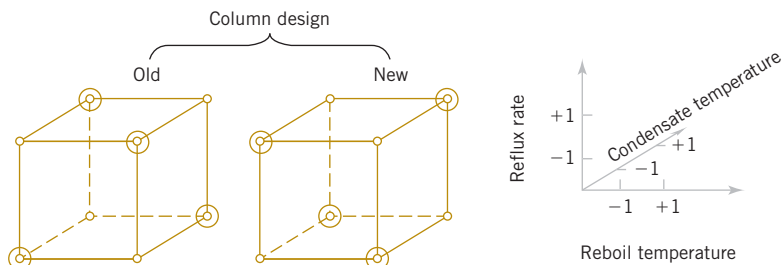
Figure 1-8 is a dot diagram of acetone concentration readings taken hourly from the distillation column described in Section 1-2.2. The large variation displayed on the dot diagram indicates considerable variability in the concentration, but the chart does not help explain the reason for the variation. 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 deviation 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.

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



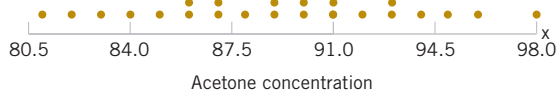


FIGURE 1-8 The dot diagram illustrates variation but does not identify the problem.

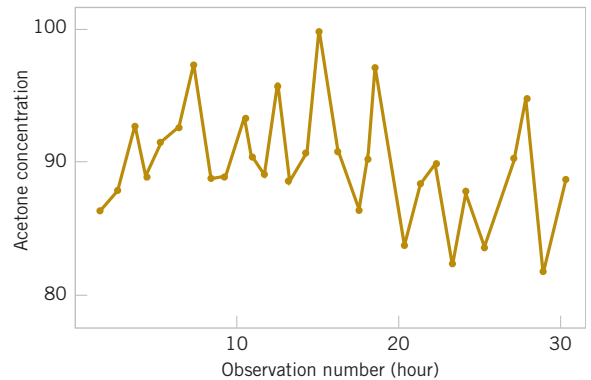


FIGURE 1-9 A time series plot of concentration provides more information than the dot diagram.

This interesting experiment points out that adjustments to a process based on random disturbances 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 lessons 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 measurements 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 understanding 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 ($\bar{x} = 91.5 \text{ g/l}$) when the process is stable. The **upper control limit** and the **lower control limit** are a pair of statistically 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.

