with website

Fourth Edition

Statistics for Health Care Management and Administration

Working with Excel

David A. Rosenthal • John F. Kros

ied of all Diseases in the Year 1592[2].

	Total	Pla.
ch 17	230	3
ch 24	351	31
ch 31	219	29
il 7	307	27
il 14	203	33
il 21	290	37
il 28	310	41
' 5	350	29
' 12	339	38
' 19	300	42
' 26	450	58
: 2	410	62
: 9	441	81
: 16	399	99
: 23	401	108
: 30	850	118
7	1440	927
14	1510	893
21	1491	258
28	1507	852
ust 4	1503	983
ust 11	1550	797
ust 18	1532	651
ust 25	1508	449
ember 1	1490	507
ember 8	1210	563
ember 15	621	451
ember 22	629	349
ember 29	450	330
ber 6	408	327
ber 13	422	323
ber 20	330	308
ber 27	320	302
ember 3	310	301
ember 10	309	209
ember 17	301	107
ember 24	321	93
ember 1	349	94
ember 8	331	86
ember 15	329	71
ember 22	386	39

Total of all that have en buried is 25886
reof of the Plague 11503

Buried of all Diseases in the Year 1603[3].

	Total	Plag.
March 17	108	3
24	60	2
31	78	6
April 7	66	4
14	79	4
21	98	8
28	109	10
May 5	90	11
12	112	18
19	122	22
26	122	32
June 2	114	30
9	131	43
16	144	59
23	182	72
30	267	158
July 7	445	263
14	612	424
The Out-Parishes this Week were joyned with the City		
July 21	1186	917
28	1728	1396
August 4	2256	1922
11	2077	1745
18	3054	2713
25	2853	2539
September 1	3385	3035
8	3078	2724
15	3129	2818
22	2456	2195
29	1961	1732
October 6	1831	1641
13	1312	1149
20	766	642
27	625	508
November 3	737	594
10	545	442
17	384	251
24	198	105
December 1	223	102
8	163	55
15	200	96
22	168	74

The Total of all is 37294

Buried of all Diseases in the Year 1625[4].

	Total	Plag.
March 17	262	4
24	226	8
31	243	11
April 7	239	10
14	256	24
21	230	25
28	305	26
May 5	292	30
12	'232	45
19	379	71
26	401	78
June 2	395	69
9	434	91
16	510	161
23	640	239
30	942	390
July 7	1222	593
14	'1781	1004
21	2850	1819
28	3583	2471
August 4	4517	3659
11	4855	4115
18	5205	4463
25	4841	4218
September 1	3897	3344
8	3157	2550
15	2148	1672
22	1994	1551
29	'236	852
October 6	833	538
13	815	511
20	651	331
27	375	134
November 3	357	89
10	319	92
17	274	48
24	231	27
December 1	190	15
8	181	15
15	168	6
22	157	1

The Total of all is 51578
Whereof of the Plague 35403

Buried of all Diseases in the Year 1630[5].

	Tot.	Pl.
June 24	205	19
July 1	209	25
8	217	43
15	250	50
22	229	40
29	279	77
August 5	250	56
12	246	65
19	269	54
26	270	67
September 2	230	66
9	259	63
16	264	68
23	274	57
30	269	56
October 7	236	66
14	261	73
21	248	60
28	214	34
November 4	242	29
11	215	29
18	200	18
25	226	7
December 2	221	20
9	198	19
16	212	5

Buried in the 97 Parishes without the walls 2696
Whereof of the Plague 190
Buried in the 16 Parishes without the walls 4813
Whereof of the Pl. 603
Buried in the 9 Out-Parishes in *Middlesex* and *Surrey*, and at the *Pest-house* 3045
Whereof of the Pl. 524
Buried in *Westminster* 566
Whereof of the Pl. 31

The Total of all the Burials this time 10545

Buried of all Diseases in the Year 1636[6].

	Total	Pla.
April 7	119	2
14	205	4
This Week these Parishes were added: S. Marg. Westminster, Lambeth Parish, S. Mary Newington, Redriff Parish, S. Mary Islington, Stepney and Hackney Parishes.		
April 21	285	14
28	259	17
May 5	251	10
12	308	55
19	299	35
26	330	62
June 2	339	77
9	345	87
16	381	103
23	304	79
30	352	104
July 7	215	81
14	372	104
21	365	120
28	423	151
August 4	491	206
11	538	283
18	638	321
25	787	429
September 1	1011	638
8	1069	650
15	1306	865
22	1229	775
29	1403	928
October 6	1405	921
13	1302	792
20	1002	555
27	900	458
November 3	1300	838
10	1104	715
17	950	573
24	857	521
December 1	614	321
8	459	167
15	385	85

The Total of the Burials this year is 23359

Buried of all Diseases in the Year 1665[7].

	Total	Plag.
December 27	291	
January 3	349	
10	394	
17	415	
24	474	
31	409	
February 7	393	
14	461	1
21	393	
28	396	
March 7	441	
14	433	
21	365	
28	353	
April 4	344	
11	382	
18	344	
25	390	2
May 2	388	
9	347	9
16	353	3
23	385	14
30	399	17
June 6	405	43
13	558	112
20	611	168
27	684	267
July 4	1006	470
11	1268	727
18	1761	1089
25	2785	1843
August 1	3014	2010
8	4030	2817
15	5319	3880
22	5568	4237
29	7496	6102
September 5	8252	6988
12	7690	6544
19	8297	7165
26	6460	5533
October 3	5068	4329
10	5068	4327
17	3219	2665
24	1806	1421
31	1388	1031
November 7	1787	1414

JOSSEY-BASS
A Wiley Brand

STATISTICS FOR HEALTH CARE MANAGEMENT AND ADMINISTRATION

STATISTICS FOR HEALTH CARE MANAGEMENT AND ADMINISTRATION

WORKING WITH EXCEL

Fourth Edition

David A. Rosenthal
John F. Kros

JOSSEY-BASS™
A Wiley Brand

Copyright © 2023 by John Wiley & Sons, Inc. All rights reserved.

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

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 Section 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, (978) 750-8400, fax (978) 750-4470, or on the web at 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, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data applied for:

Paperback ISBN: 9781119901679

Cover image: © Table in Economic Writings (vol. 2) by William Petty (1899), djvu p. 121
Author Photos: Courtesy of the Authors
Cover design: Wiley

CONTENTS

Preface xiii
Introducing Excel xiii
So How Did We Get to Here? xiii
Intended Level of the Textbook xiv
Textbook Organization xiv
Leading by Example(s) xv
Acknowledgments xvii
The Authors xix
About the Companion Website xxi

Part 1 **1**

Chapter 1 Statistics and Excel 3

1.1 How This Book Differs from Other Statistics Texts 3
1.2 Statistical Applications in Health Policy and Health Administration 4
Exercises for Section 1.2 14
1.3 What Is the "Big Picture"? 15
1.4 Some Initial Definitions 16
Exercises for Section 1.4 26
1.5 Five Statistical Tests 28
Exercises for Section 1.5 30

Chapter 2 Excel as a Statistical Tool 33

2.1 The Basics 33
Exercises for Section 2.1 35
2.2 Working and Moving Around in a Spreadsheet 36
Exercises for Section 2.2 41
2.3 Excel Functions 41
Exercises for Section 2.3 46
2.4 The `=IF()` Function 47
Exercises for Section 2.4 50
2.5 Excel Graphs 51
Exercises for Section 2.5 56

2.6 Sorting a String of Data . . . 57
Exercise for Section 2.6 . . . 60
2.7 The Data Analysis Pack . . . 61
2.8 Functions That Give Results in More than One Cell . . . 63
Exercises for Section 2.8 . . . 66
2.9 The Dollar Sign ($) Convention for Cell References . . . 67

Chapter 3 Data Acquisition: Sampling and Data Preparation . . . 71

3.1 The Nature of Data . . . 71
Exercises for Section 3.1 . . . 78
3.2 Sampling . . . 79
Exercises for Section 3.2 . . . 93
3.3 Data Access and Preparation . . . 94
Exercises for Section 3.3 . . . 107
3.4 Missing Data . . . 108

Chapter 4 Data Display: Descriptive Presentation, Excel Graphing Capability . . . 111

4.1 Creating, Displaying, and Understanding Frequency Distributions . . . 111
Exercises for Section 4.1 . . . 129
4.2 Using the Pivot Table to Generate Frequencies of Categorical Variables . . . 131
Exercises for Section 4.2 . . . 135
4.3 A Logical Extension of the Pivot Table: Two Variables . . . 135
Exercises for Section 4.3 . . . 140

Chapter 5 Basic Concepts of Probability . . . 141

5.1 Some Initial Concepts and Definitions . . . 141
Exercises for Section 5.1 . . . 150
5.2 Marginal Probabilities, Joint Probabilities, and Conditional Probabilities . . . 150
Exercises for Section 5.2 . . . 160
5.3 Binomial Probability . . . 161
Exercises for Section 5.3 . . . 171
5.4 The Poisson Distribution . . . 173
Exercises for Section 5.4 . . . 178
5.5 The Normal Distribution . . . 178

Chapter 6 Measures of Central Tendency and Dispersion: Data Distributions 183

6.1 Measures of Central Tendency and Dispersion 183
Exercises for Section 6.1 196
6.2 The Distribution of Frequencies 197
Exercises for Section 6.2 208
6.3 The Sampling Distribution of the Mean 209
Exercises for Section 6.3 219
6.4 Mean and Standard Deviation of a Discrete Numerical Variable 220
Exercises for Section 6.4 222
6.5 The Distribution of a Proportion 222
Exercises for Section 6.5 227
6.6 The *t* Distribution 227
Exercises for Section 6.6 232

Part 2 235

Chapter 7 Confidence Limits and Hypothesis Testing 237

7.1 What Is a Confidence Interval? 237
Exercises for Section 7.1 243
7.2 Calculating Confidence Limits for Multiple Samples . . . 244
Exercises for Section 7.2 246
7.3 What Is Hypothesis Testing? 247
Exercises for Section 7.3 249
7.4 Type I and Type II Errors 250
Exercises for Section 7.4 266
7.5 Selecting Sample Sizes 267
Exercises for Section 7.5 269

Chapter 8 Statistical Tests for Categorical Data 271

8.1 Independence of Two Variables 271
Exercises for Section 8.1 282
8.2 Examples of Chi-Square Analyses 283
Exercises for Section 8.2 289
8.3 Small Expected Values in Cells 290
Exercises for Section 8.3 292

Chapter 9 *t* Tests for Related and Unrelated Data 295

9.1 What Is a *t* Test? 295
Exercises for Section 9.1 302
9.2 A *t* Test for Comparing Two Groups 303
Exercises for Section 9.2 316
9.3 A *t* Test for Related Data 318
Exercises for Section 9.3 321

Chapter 10 Analysis of Variance 323

10.1 One-Way Analysis of Variance 323
Exercises for Section 10.1 339
10.2 ANOVA for Repeated Measures 340
Exercises for Section 10.2 348
10.3 Factorial Analysis of Variance 349
Exercises for Section 10.3 362

Chapter 11 Simple Linear Regression 365

11.1 Meaning and Calculation of Linear Regression 365
Exercises for Section 11.1 373
11.2 Testing the Hypothesis of Independence 374
Exercises for Section 11.2 380
11.3 The Excel Regression Add-In 381
Exercises for Section 11.3 388
11.4 The Importance of Examining the Scatterplot 388
11.5 The Relationship between Regression and the *t* Test . . . 391
Exercises for Section 11.5 392

Chapter 12 Multiple Regression: Concepts and Calculation 395

12.1 Introduction 395
Exercises for Section 12.1 406

Chapter 13 Extensions of Multiple Regression 409

13.1 Dummy Variables in Multiple Regression 409
Exercises for Section 13.1 420
13.2 The Best Regression Model 421
Exercises for Section 13.2 431
13.3 Correlation and Multicolinearity 432
Exercises for Section 13.3 435
13.4 Nonlinear Relationships 435
Exercises for Section 13.4 447

Chapter 14 Analysis with a Dichotomous Categorical Dependent Variable 449

14.1 Introduction to the Dichotomous Dependent Variable . . . 450
14.2 An Example with a Dichotomous Dependent Variable: Traditional Treatments 451
Exercises for Section 14.2 462
14.3 Logit for Estimating Dichotomous Dependent Variables . . 463
Exercises for Section 14.3 475
14.4 A Comparison of Ordinary Least Squares, Weighted Least Squares, and Logit 476
Exercises for Section 14.4 480

Appendix A Multiple Regression and Matrices 481

An Introduction to Matrix Math 481
Addition and Subtraction of Matrices 482
Multiplication of Matrices 483
Matrix Multiplication and Scalars 484
Finding the Determinant of a Matrix 484
Matrix Capabilities of Excel 486
Explanation of Excel Output Displayed with Scientific Notation . . 489
Using the *b* Coefficients to Generate Regression Results 490
Calculation of All Multiple Regression Results 491
Exercises for Appendix A 494

References 497
Glossary 499
Index 513

David dedicates this edition to all those who have been called to a career in health care, and who serve others each day with passion, patience, and a commitment to excellence.

—David

John dedicates this edition to his entire family (grandmothers, grandfathers, mother, father, sisters, brother, wife, and children), who have always supported him in all he does, inspired him to always strive for excellence, and loved him all these years.

—John

PREFACE

The study and use of statistics have come a long way since the advent of computers. Particularly, computers have reduced both the effort and the time involved in the statistical analysis of data. But this ease of use has been accompanied by some difficulties. As computers became more and more proficient at carrying out statistical operations of increasing complexity, the actual operations—and what they actually meant and did—became more and more distant from the user. It became possible to do a wide variety of statistical operations with a few lines or words of commands to the computer. But the average student, even the average serious user of statistics, found the increasingly complex operations increasingly difficult to access and understand.

Introducing Excel

Sometime in the late 1980s, Microsoft Excel became available, and with it came the ability to carry out a wide range of statistical operations—and to understand the operations that were being carried out—in a spreadsheet format. John's first introduction to Excel was a revelation. It came during his MBA studies and continued through his doctoral studies and even in his first industry job. In fact, John quickly became somewhat indispensable in that first industry job for the plain fact that he was the most proficient of his peers at Excel. Through the years he found himself using Excel to complete all kinds of tasks (since he was too stubborn to learn to program properly). He discovered that Excel was not only a powerful statistical tool but also, more important, a powerful learning tool. When he began to teach the introductory course in business decision modeling to MBA students, Excel seemed to him to be the obvious medium for the course.

So How Did We Get to Here?

At the time John started using Excel in his teaching, there were a few textbooks devoted to statistics using Excel. However, none fit his needs very well, so he wrote *Spreadsheet Modeling for Business Decision Modeling.*

That was about the time John met David. David had earned his doctorate in technology management and had worked in the health care industry for more than 10 years (which ensures that the health care–specific examples and scenarios used in this book are appropriate). He discovered the power of Excel's statistical analysis functionality by using it to calculate the multiple regression and correlation analysis required for his doctoral dissertation.

Through his friend, Scott Bankard, John learned that the author of a successful text in the use of Excel to solve statistical problems in health care administration was looking for someone to revise that text. In turn, John and David became the coauthors of the revised text.

Intended Level of the Textbook

The original text was designed as an introductory statistics text for students at the advanced undergraduate level or for a first course in statistics at the master's degree level. It was intended to stand alone as the book for the only course a student might have in statistics. The same is true for the revised text, which includes some enhancements and updates that provide a good foundation for more advanced courses as well. Furthermore, since the book relies on Excel for all the calculations of the statistical applications, it was also designed to provide a statistical reference for people working in the health field who may have access to Excel but not to other dedicated statistical software. This is valuable in that a copy of Excel resides on the PC of almost every health care professional. Further, no additional appropriations would have to be made for proprietary software and there would be no wait for the "stat folks."

Textbook Organization

The revised edition of the text has been updated for use with the latest version of Microsoft Office Excel. It provides succinct instruction in the most commonly used techniques and shows how these tools can be implemented using the most current version of Excel for Windows. The revised text also focuses on developing both algebraic and spreadsheet modeling skills. Algebraic formulation and spreadsheets are juxtaposed to help develop conceptual thinking skills. Step-by-step instructions in Excel and numerous annotated screenshots make examples easy to follow and understand. Emphasis is placed on the model formulation and interpretation rather than on computer code or algorithms.

The book is organized into two major parts: Part 1, Chapters 1 through 6, presents Excel as a statistical tool and discusses hypothesis testing. Part 1 introduces the use of statistics in health policy and health

administration–related fields, Excel as a statistical tool, data preparation and the data display capabilities of Excel, and probability, the foundation of statistical analysis. For students and other users of the book truly familiar with Excel, much of the material in Chapter 2, Chapter 3, and Chapter 4, particularly, could be covered very quickly.

Part 2, which includes Chapters 7 through 14, is devoted to the subject of hypothesis testing, the basic function of statistical analysis. Chapter 7 provides a general introduction to the concept of hypothesis testing. Each subsequent chapter provides a description of the major hypothesis testing tool for a specific type of data. Chapter 8 discusses the use of the chi-square statistic for assessing data for which both the independent and dependent variables are categorical. Chapter 9, on *t* tests, discusses the use of the *t* test for assessing data in which the independent variable is a two-level categorical variable and the dependent variable is a numerical variable. Chapter 10 is devoted to analysis of variance, which provides an analytical tool for a multilevel categorical independent variable and a numerical dependent variable. Chapters 11 through 13 are devoted to several aspects of regression analysis, which deals with numerical variables both as independent and dependent variables. Finally, Chapter 14 deals with numerical independent variables and dependent variables that are categorical and take on only two levels and introduces the use of Logit.

Leading by Example(s)

Each chapter of the book is structured around examples demonstrated extensively with the use of Excel displays. The chapters are divided into sections, most of which include step-by-step discussions of how statistical problems are solved using Excel, including the Excel formulae. Each section in a chapter is followed by exercises that address the material covered in that section. Most of these exercises include the replication of examples from that section. The purpose is to provide students an immediate reference with which to compare their work and determine whether they are able to correctly carry out the procedure involved. Additional exercises are provided on the same subjects for further practice and to reinforce the learning gained from the section. Data for all the exercises are included on the web at **www.wiley.com/go/kros4e**, and may be accessed by file references given in the examples themselves. Additional materials, such as videos, podcasts, and readings, can be found at www.josseybasspublichealth.com.

A supplemental package available to instructors includes all answers to the section exercises. In addition, the supplemental package will contain exam questions with answers and selected Excel spreadsheets that can be

used for class presentations, along with suggestions for presenting these materials in a classroom. However, the book can be effectively used for teaching without the additional supplemental material.

Users who would like to provide feedback, suggestions, corrections, examples of applications, or whatever else can e-mail me at krosj@ecu.edu.

Please feel free to contact me and provide any comments you feel are appropriate.

ACKNOWLEDGMENTS

As always this newly revised version of the text would not have been possible without the support and guidance of numerous colleagues, friends, family, and all those poor souls who had to listen to us bounce ideas off of them, or for that matter anyone that just had to listen to us!

David thanks his wife Allyson and their entire extended family for their steadfast love and support. David also thanks his friend and colleague John for the years of collaboration and hard work pulling together the various editions of this textbook. It has been an honor and a privilege.

John thanks his wife, Novine, and his daughters, Samantha and Sabrina, for always being by his side and encouraging him in the special way they do when the light at the end of the tunnel starts to dim. Samantha always reminds her dad that she loves him, and maybe someday he will be cool again (no date has been set yet). Sabrina tells her dad that she expects great things from the East Carolina University Pirate football squad, the Texas Longhorn and Nebraska Cornhusker football teams, and the Virginia Cavaliers even if she is a Wolfpacker now. She also invites anyone interested to eat brunch on any given Sunday at her favorite establishment, the West End Dining Hall on East Carolina University's campus. John would like to say thank you to Novine, Samantha, and Sabrina and that he loves them very much. John also must thank his parents, Bernie and Kaye, who have always supported him, even when they didn't exactly know what he was writing about. Finally, John has to thank Scott Bankard for setting things in motion way back in 2007 and suggesting the project.

—David A. Rosenthal, PhD, and John F. Kros, PhD

THE AUTHORS

David A. Rosenthal is a Chief Information Officer with Campus Works, Inc., a higher education consulting firm that collaborates with colleges and universities to develop and implement insightful strategies that improve institutional effectiveness and enhance the student experience. He has over 20 years of health care industry experience in both academic and practitioner settings, having served in roles specific to hospital information technology leadership, multispecialty practice administration, and ambulatory services project management. Rosenthal earned a Master of Public Administration degree from Valdosta State University, and a PhD in Technology Management from Indiana State University. He is a proud veteran of the U.S. Air Force, and a Permanent Deacon in the Roman Catholic Church. David and his wife Allyson reside in Oakland, Tennessee.

John F. Kros is the Vincent K. McMahon Distinguished Professor in the Marketing and Supply Chain Management Department in the College of Business at East Carolina University, in Greenville, North Carolina. He teaches business decision modeling, statistics, operations and supply chain management, and purchasing and materials management courses. Kros was honored as the College of Business's Scholar/Teacher for 2004–2005, again in 2009–2010, was awarded the College of Business Commerce Club's highest honor, the Teaching Excellence Award, for 2006

and again in 2011, in 2013–2014 was awarded the East Carolina Alumni Association Outstanding Teaching Award, and in 2014–2015 was awarded the Board of Governors Distinguished Professor for Teaching Award. Kros earned his PhD in systems engineering from the University of Virginia, his MBA from Santa Clara University, and his BBA from the University of Texas at Austin. His research interests include health care operations, applied statistics, design of experiments, multi-objective decision making, Taguchi methods, and applied decision analysis. His textbook titled *Spreadsheet Modeling for Business Decisions* is in the fifth edition. He is also coauthor of *Health Care Operations and Supply Chain Management*. He enjoys spending his free time with his beautiful red-headed wife, Novine, and their two beautiful daughters, Samantha and Sabrina, traveling, snow skiing, vegetable gardening, spending time with his family and old fraternity brothers, watching college football, and attempting to locate establishments that provide quality food and liquid refreshment.

ABOUT THE COMPANION WEBSITE

This book is accompanied by a companion website:

www.wiley.com/go/kros4e

The website includes:

- Datasets
- PPTs
- Excel sheets
- Test banks

You will need a password to access the instructor material.

PART 1

1. Statistics and Excel
2. Excel as a Statistical Tool
3. Data Acquisition: Sampling and Data Preparation
4. Data Display: Descriptive Presentation, Excel Graphing Capability
5. Basic Concepts of Probability
6. Measures of Central Tendency and Dispersion: Data Distributions

CHAPTER 1

STATISTICS AND EXCEL

The statistics on sanity are that one out of every four Americans is suffering from some form of mental illness. Think of your three best friends. If they're okay, then it's you.

—Rita Mae Brown

Statistics is a subject that for many people is pure tedium. For others, it is more likely to be anathema. Still others find statistics interesting, even stimulating, but they are usually in the minority in any group.

This book is premised on the recognition that in the health care industry, as indeed among people in any industry or discipline, there are at least these three different views of statistics, and that any statistics class is likely to be made up more of the first two groups than the last one. This book provides an introduction to statistics in health policy and administration that is relevant, useful, challenging, and informative.

1.1 How This Book Differs from Other Statistics Texts

The primary difference between this statistics text and most others is that this text uses Microsoft Excel as the tool for carrying out statistical operations and understanding statistical concepts as they relate to health policy and health administration issues. This is not to say that no other statistics texts use Excel. Levine, Stephan, Szabat (2013) have produced a very usable text, *Statistics for Managers Using Microsoft Excel*. But their book focuses almost exclusively on non–health-related topics. We agree that the closer

LEARNING OBJECTIVES

- Understand how this book differs from other statistics texts
- Understand how knowledge of statistics may be beneficial to health policy or health administration professionals
- Understand the "big picture" with regard to the use of statistics for health policy and administration
- Understand the definitions of the following terms:
  - Populations and samples
  - Random and nonrandom samples
  - Types of random samples
  - Variables, independent and dependent
- Identify the five separate statistical tests: chi-square test, the *t* test, analysis of variance (ANOVA), regression analysis, and Logit

the applications of statistics are to students' real-life interests and experiences, the more effective students will be in understanding and using statistics. Consequently, this book focuses its examples entirely on subjects that should be immediately familiar to people in the health care industry.

Excel, which most people know as a *spreadsheet* program for creating budgets, comparing budgeted and expended amounts, and generally fulfilling accounting needs, is also a very powerful statistical tool. Books that do not use Excel for teaching statistics generally leave the question of how to carry out the actual statistical operations in the hands of the student or the instructor. It is often assumed that relatively simple calculations, such as means, standard deviations, and *t* tests, will be carried out on paper or with a calculator. For more complicated calculations, the assumption is usually that a dedicated statistical package, such as SAS, SPSS, STATA, or SYSTAT, will be used. There are at least two problems with this approach that we hope to overcome in this book. First, calculations done on paper, or even those done with a calculator, can make even simple statistical operations overly tedious and prone to errors in arithmetic. Second, because dedicated statistical packages are designed for use rather than for teaching, they often obscure the actual process of calculating the statistical results, thereby hindering students' understanding of both how the statistic is calculated and what the statistic means.

In general, this is not true of Excel. It is true that when using this book, a certain amount of time must be devoted to the understanding of how to use Excel as a statistical tool. But once that has been done, Excel makes the process of carrying out the statistical procedures under consideration relatively clear and transparent. The student should end up with a better understanding of what the statistic means, through an understanding of how it is calculated, and not simply come away with the ability to get a result by entering a few commands into a statistical package. This is not to say that Excel cannot be used to eliminate many of the steps needed to get particular statistical results. A number of statistical tests and procedures are available as add-ins to Excel. However, using Excel as a relatively powerful—yet transparent—calculator can lead to a much clearer understanding of what a statistic means and how it may be used.

1.2 Statistical Applications in Health Policy and Health Administration

When teaching statistics to health policy and health administration students, we often encounter the same question: "All these statistics are fine, but how do they apply to anything I am concerned with?" The question not only is a reasonable one, but also points directly to one of the most

important and difficult challenges for a statistics teacher, a statistics class, or a statistics text. How can it be demonstrated that these statistics have any real relevance to anything that the average person working in the health care industry ever needs to know or do?

To work toward a better understanding of why and when the knowledge of statistics may be useful to someone working in health policy or health administration, we've selected six examples of situations in which statistical applications can play a role. All six of these examples were inspired by real problems faced by students in statistics classes, and they represent real statistical challenges that students have faced and hoped to solve. In virtually every case, the person who presented the problem recognized it as one that could probably be dealt with using some statistical tool. But also in every case, the solution to the problem was not obvious in the absence of some understanding of statistics. Although these case examples are not likely to resonate with every reader, perhaps they will give many readers a little better insight into why knowledge of statistics can be useful.

Documentation of Medicare Reimbursement Claims

The Pentad Home Health Agency provides home health services in five counties of an eastern state. The agency must be certain that its *Medicare* reimbursement claims are appropriately and correctly documented in order to ensure that Medicare will process these claims and issue benefits in a timely manner. All physician orders, including medications, home visits for physical therapy, home visits of skilled nursing staff, and any other orders for service, must be correctly documented on a Form CMS-485. Poorly or otherwise inadequately prepared documentation can lead to rejection or delay in processing of the claim for reimbursement by the Centers for Medicare and Medicaid Services (CMS).

Pentad serves about 800 clients in the five-county region. In order to assure themselves that all records are properly documented, the administration runs a chart audit of 1 in 10 charts each quarter. The audit seeks to determine (1) whether all orders indicated in the chart have been carried out and (2) if the orders have been correctly documented in the Form CMS-485. Orders that have not been carried out, or orders incorrectly documented, lead to follow-up training and intervention to address these issues and ensure that the orders and documentation are properly prepared going forward.

Historically, the chart audit has been done by selecting each tenth chart, commencing at the beginning or at the end of the chart list. Typically, the chart audit determines that the majority of charts, usually 85 to 95 percent, have been correctly documented. But there are occasionally areas, such

as in skilled nursing care, where the percentage of correct documentation may fall below that level. When this happens, the administration initiates appropriate corrective action.

Sampling, Data Display, and Probability

One of the questions of the audit has been the selection of the sample. Because the list of clients changes relatively slowly, the selection of every tenth chart often results in the same charts being selected for audit from one quarter to the next. That being the case, a different strategy for chart selection is desirable. It has been suggested by statisticians that using a strictly random sample of the charts might be a better way to select them for quarterly review, as this selection would have a lesser likelihood of resulting in a review of the same charts from quarter to quarter. But how does one go about drawing a strictly random sample from any population? Or, for that matter, what does "strictly random" actually mean and why is it important beyond the likelihood that the same files may not be picked from quarter to quarter? These questions are addressed by statistics, specifically the statistics associated with sample selection and data collection.

Another question related to the audit concerns when to initiate corrective action. Suppose a sample of 1 in 10 records is drawn (for 800 clients that would be 80 records) and it is discovered that 20 of the records have been incorrectly documented. Twenty of 80 records incorrectly documented would mean that only 75 percent of the records were correctly documented. This would suggest that an intervention should be initiated to correct the documentation problem. But it was a *sample* of the 800 records that was examined, not the entire 800. Suppose that the 20 incorrectly documented records were, by the luck of the draw, so to speak, the only incorrectly documented records in the entire 800. That would mean that only 2.5 percent of the cases were incorrectly documented.

If the required corrective action were an expensive five-day workshop on correct documentation, the agency might not want to incur that expense when 97.5 percent of all cases are correctly documented. But how would the agency know from a sample what proportion of the total 800 cases might be incorrectly documented, and how would they know the likelihood that fewer than, say, 85 percent of all cases were correctly documented if 75 percent of a sample were correctly documented? This, again, is a subject of statistics.

Emergency Trauma Color Code

The emergency department (ED) of a university hospital was the site of difficulties arising from poor response time to serious trauma. Guidelines