

The background of the book cover features a green grid with dashed lines. A yellow line graph is overlaid on the grid, showing a fluctuating trend with a prominent peak. The text is presented in white and yellow on dark green and black rectangular backgrounds.

Third Edition

Statistics for Health Care Management and Administration

Working with Excel

JOSSEY-BASS™
A Wiley Brand

John F. Kros • David A. Rosenthal

**STATISTICS FOR HEALTH CARE MANAGEMENT
AND ADMINISTRATION**

STATISTICS FOR HEALTH CARE MANAGEMENT AND ADMINISTRATION

WORKING WITH EXCEL

Third Edition

**John F. Kros
David A. Rosenthal**

JB JOSSEY-BASS™
A Wiley Brand

Copyright ©2016 by John Wiley & Sons, Inc. All rights reserved.

Published by Jossey-Bass

A Wiley Brand

One Montgomery Street, Suite 1000, San Francisco, CA 94104-4594—www.josseybass.com

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 978-646-8600, 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 www.wiley.com/go/permissions.

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. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Readers should be aware that Internet Web sites offered as citations and/or sources for further information may have changed or disappeared between the time this was written and when it is read.

Jossey-Bass books and products are available through most bookstores. To contact Jossey-Bass directly call our Customer Care Department within the U.S. at 800-956-7739, outside the U.S. at 317-572-3986, or fax 317-572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Names: Kros, John F. | Rosenthal, David A. | Veney, James E. Statistics for health care professionals working with Excel.

Title: Statistics for health care management and administration : working with Excel / John F. Kros, David A. Rosenthal.

Other titles: Statistics for health care professionals working with Excel.

Description: Third edition. | Hoboken : Wiley, 2016. | Series: Public health/epidemiology and biostatistics | Revised edition of Statistics for health care professionals working with Excel, 2009. | Includes index.

Identifiers: LCCN 2015019360 (print) | LCCN 2015039488 (ebook) | ISBN 9781118712658 (paperback) | ISBN 9781118712641 (pdf) | ISBN 9781118712764 (epub)

Subjects: LCSH: Medical statistics. | Microsoft Excel (Computer file) | MESH: Microsoft Excel (Computer file) | Data Interpretation, Statistical. | Health Services Administration. | Automatic Data Processing. | Health Policy. | BISAC: MEDICAL / Public Health.

Classification: LCC RA409.5 .K76 2016 (print) | LCC RA409.5 (ebook) | DDC 610.2/1—dc23
LC record available at <http://lccn.loc.gov/2015019360>

Cover design: Wiley

Cover image: © Fariz Alikishibayov/Shutterstock; Peter Nadolski/Shutterstock; autsawinuttisin/Shutterstock

Printed in the United States of America

The second edition was published under the title *Statistics for Health Care Professionals: Working With Excel*.

THIRD EDITION

HB Printing 10 9 8 7 6 5 4 3 2 1

PB Printing 10 9 8 7 6 5 4 3 2 1

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
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	<i>t</i> Tests for Related and Unrelated Data	295
9.1	What Is a <i>t</i> Test?	295
	Exercises for Section 9.1	302
9.2	A <i>t</i> Test for Comparing Two Groups	303
	Exercises for Section 9.2	316
9.3	A <i>t</i> 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 <i>t</i> 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 Multicollinearity	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 Variable449
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 Matrices481
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

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

David dedicates this edition to his grandchildren—Addison and Grayson—who will always have his heart, and to the memory of his mother and father, who instilled in him an appreciation for the blessings of a loving family.

—David

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 Microsoft Office Excel 2013. 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 2013 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/kros3e, 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!

Many people contributed to this book as it now appears—in particular, several faculty at the University of North Carolina who contributed to the original edition and to this revision in various ways.

We would like to thank proposal reviewers Nan Liu, Xinliang Liu, Lawrence A. West, Jr., James Porto, and Graciela E. Silva, who provided valuable feedback on the original book proposal. Echu Liu, John P. Gaze, and James Porto provided thoughtful and constructive comments on the complete draft manuscript.

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. At present Samantha is 13 years old and always reminds her dad that she loves him, and maybe someday he will be cool again (no date has been set yet). Sabrina is 11 years old and 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. 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 of Vidant Health for setting things in motion way back in 2007 and suggesting the project.

David thanks his wife Allyson for her unending and unwavering support, and his son, daughter, and son-in-law for being so easily impressed by his modest professional achievements.

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

THE AUTHORS

John F. Kros is the Vincent K. McMahon Distinguished Professor of Business 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 logistics and materials management courses. Kros was honored as the College of Business's Scholar/Teacher for 2004–2005, again in 2009–2010, and was awarded the College of Business Commerce Club's highest honor, the Teaching Excellence Award, for 2006 and again in 2011. 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. In 2014, the fourth edition of his textbook titled *Spreadsheet Modeling for Business Decisions* was printed. He is also coauthor of *Health Care Operations and Supply Chain Management*, published in 2013. 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.

David A. Rosenthal is a professor and chair of Health Care Management at Baptist College of Health Sciences in Memphis, Tennessee. Rosenthal earned a master of public administration degree from Valdosta State University in 1996, and a PhD in technology management from Indiana State University in 2002. He has over 20 years of health care experience in both academic and practitioner settings, having served in roles specific to health care information technology leadership, multispecialty practice management, and ambulatory services project management. Rosenthal served for two years as director of the state of Tennessee's Health Information Exchange (HIE) Evaluation Project while a faculty member in the division of Health Systems Management and Policy in the School of Public

Health at the University of Memphis from 2009 to 2013. He also served as director of Statewide eHealth Initiatives while a faculty member in the Department of Health Informatics and Information Management at the University of Tennessee Health Science Center in Memphis from 2007 to 2009. Rosenthal currently resides in Memphis, Tennessee, with his wife, Allyson, and their extended family.

**STATISTICS FOR HEALTH CARE MANAGEMENT
AND ADMINISTRATION**

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

STATISTICS AND EXCEL

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 useable text, *Statistics for Managers Using Microsoft Excel*. But their book focuses almost exclusively on non–health-related topics. We agree that the closer 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.

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*

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