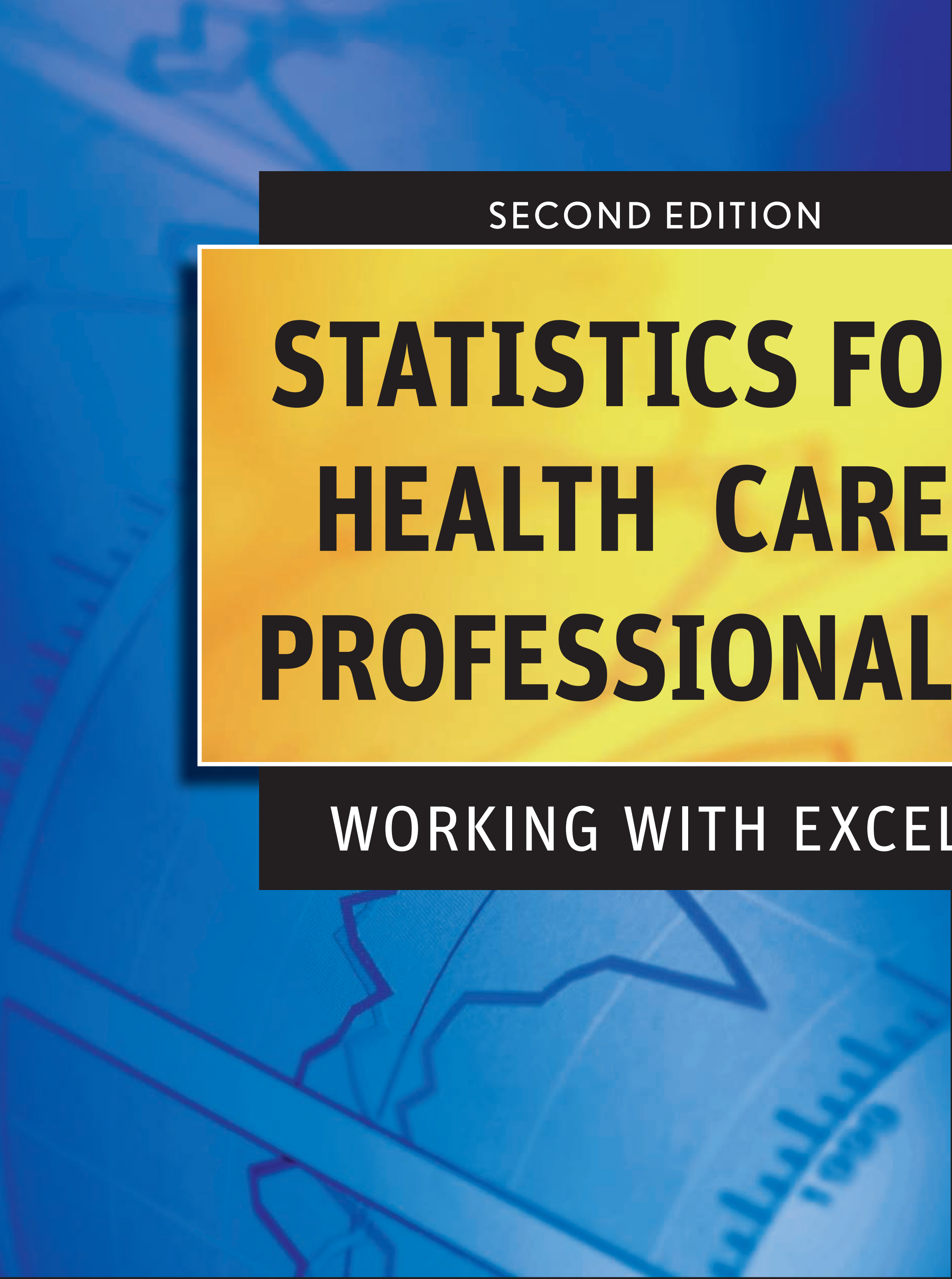

SECOND EDITION

STATISTICS FOR HEALTH CARE PROFESSIONALS

WORKING WITH EXCEL

JAMES E. VENEY • JOHN F. KROS • DAVID A. ROSENTHAL

STATISTICS FOR HEALTH CARE PROFESSIONALS

STATISTICS FOR HEALTH CARE PROFESSIONALS

Working with Excel

(Second edition of Statistics for Health Policy and Administration Using Microsoft Excel*)*

JAMES E. VENEY

JOHN F. KROS

DAVID A. ROSENTHAL

JOSSEY-BASS
A Wiley Imprint
www.josseybass.com

Published by Jossey-Bass
A Wiley Imprint
989 Market Street, San Francisco, CA 94103-1741—www.josseybass.com

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.

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.

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.

Jossey-Bass also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Veney, James E.
Statistics for health care professionals working with Excel / James E. Veney, John F. Kros, David A. Rosenthal. — 1st ed.
p. ; cm.
Rev. ed. of: Statistics for health policy and administration using Microsoft Excel / James E. Veney. c2003.
Includes bibliographical references and index.
ISBN 978-0-470-39331-4 (pbk.)
1. Medical statistics. 2. Microsoft Excel (Computer file) I. Kros, John F. II. Rosenthal, David A., 1959- III. Veney, James E. Statistics for health policy and administration using Microsoft Excel. IV. Title.
[DNLM: 1. Microsoft Excel (Computer file) 2. Data Interpretation, Statistical. 3. Health Services Administration. 4. Automatic Data Processing. 5. Health Policy. W 84.1 V457s 2009]
RA409.5.V466 2009
610.72—dc22
2009017420

Printed in the United States of America
SECOND EDITION

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

CONTENTS

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. John and David approached the author, James Veney, and the three of them decided to work together on the revision.

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 and 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 2007. 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 formulations and spreadsheets are juxtaposed to help develop conceptual thinking skills. Step-by-step instructions in Excel 2007 and numerous annotated screen shots make examples easy to follow and understand. Emphasis is placed on model formulation and interpretation rather than on computer code or algorithms.

The book is organized into two major parts: Part One, chapters one through six, presents Excel as a statistical tool and discusses hypothesis testing. Part One 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 chapters two, three, and four, particularly, could be covered very quickly.

Part Two, which includes chapters seven through fourteen, is devoted to the subject of hypothesis testing, the basic function of statistical analysis. Chapter seven 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 eight discusses the use of the chi-square statistic for assessing data for which both the independent and dependent variables are categorical. Chapter nine, 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 ten is devoted to analysis of variance, which provides an analytical tool for a multilevel categorical independent variable and a numerical dependent variable. Chapters eleven through thirteen are devoted to several aspects of regression analysis, which deals with numerical variables both as independent and dependent variables. Finally, Chapter fourteen 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 http://www.josseybass.com/go/veney, and may be accessed by file references given in the examples themselves.

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. The Web site for additional resources and information is http://www.josseybass.com/go/veney2e.

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

ACKNOWLEDGMENTS

As always this effort 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 in various ways. Jim Porto deserves particular thanks as the person who helped in the early development of the first manuscript and first edition of the text. James Veney, the author of the first edition, deserves many thanks for keeping the idea moving forward and for guidance along the revision's path.

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 6 years old and always reminds John that she loves him and maybe someday he will be as cool as the Jonas Brothers. Sabrina is 5 years old and tells John that she expects great things from the East Carolina University Pirate football squad, the Texas Longhorn and Nebraska Cornhusker football teams as well as the Virginia Cavaliers. 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 and suggesting the project.

David thanks his wife Allyson for her unending and unwavering support. He is especially grateful to Beth Bowman and Marcia Sharp at the University of Tennessee Health Science Center for their contributions and edits to chapter one.

John and David thank all of those who received email attachments, Dr Pepper-stained hard copies, or random sheets of paper scribbled with half-blown ideas. We appreciate all your input or, for that matter, the simple fact you took time-out to look at our work and listen to our ramblings. And to anyone that we have forgotten to mention, thank you very much.

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

THE AUTHORS

James E. Veney is professor emeritus of health policy and administration at the University of North Carolina at Chapel Hill. He has served as director of both the master's and doctoral programs in the department and has taught courses in research methodology, evaluation methodology, statistics, and international health systems.

John F. Kros is an associate 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 logistics and materials management courses. Kros was honored as the College of Business' Scholar/Teacher for 2004–2005 and was awarded the College of Business Commerce Club's highest honor, the Teaching Excellence Award, for 2006. 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 2007, Kros published a textbook titled *Spreadsheet Modeling for Business Decisions*. 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 Rosenthal is an associate professor in the Department of Health Informatics and Information Management at the University of Tennessee's Health Science Center in Memphis, where he also provides consulting on statewide e-health and telehealth initiatives. Rosenthal received his PhD in technology management with emphasis in digital communications from Indiana State University in 2002 and has more than 20 years of information technology industry experience. His ten years of health care experience includes work in executive, managerial, and advisory roles in hospital information technology departments. Rosenthal was also a faculty member in the Department of Management Information Systems at East Carolina University where he created the Center for Healthcare Information Systems Research for the facilitation of interdisciplinary research projects. His research interests include innovation in telemedicine/telehealth technologies, the existence of the digital divide in health care, and the adoption of electronic health record technology. Rosenthal resides in Memphis, Tennessee with his wife, Allyson, and their two children.

STATISTICS FOR HEALTH CARE PROFESSIONALS

PART

1

CHAPTER

1

STATISTICS AND EXCEL

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

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, Krehbiel, and Berenson (2007) 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 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.

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 is not only a reasonable one, but it 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 CMS.

Pentad serves about eight hundred clients in the five-county region. In order to assure themselves that all records are properly documented, the administration runs a chart audit of one in ten 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 one in ten records is drawn (for eight hundred clients, that would be eighty records) and it is discovered that twenty of the records have been incorrectly documented. Twenty of eighty 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 eight hundred records that was examined, not the entire eight hundred. Suppose that the twenty incorrectly documented records were, by the luck of the draw, so to speak, the only incorrectly documented records in the entire eight hundred. 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 eight hundred 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 indicate that a trauma surgeon must attend for a certain level of trauma severity within twenty minutes and that other trauma, still severe but less so, should be attended by a trauma nurse within a comparable time. In general, it had been found that the response time for the ED in the university hospital was more or less the same for all levels of severity of trauma—too long for severe cases and often quicker than necessary, given competing priorities for less severe cases.

The ED director knew that when a trauma case was en route to the hospital, the ambulance attendants called the ED to advise that an emergency was on its way. Part of the problem as perceived by the director of the ED was that the call-in did not differentiate the trauma according to severity. The ED director decided to institute a system whereby the ambulance attendants would assign a code red to the most severe trauma cases, a code yellow to less severe trauma cases, and no color code to the least severe trauma cases. The color code of the trauma would be made known to the ED as the patient was being transported to the facility. The intent of this coding was to ensure that the most severe traumas were attended within the twenty-minute guidelines. This in turn was expected to reduce the overall time from admission to the ED to discharge of the patient to the appropriate hospital department. (All trauma cases at the red or yellow level of severity are transferred from the ED to a hospital department.)

Descriptive Statistics, Confidence Limits, and Categorical Data

A major concern of the director of the ED was whether the new system actually reduced the overall time between admission to the ED, treatment of the patient in the ED, and discharge to the appropriate hospital department. The director of the ED had considerable information about each ED admission going back a period of several months before the implementation of the new color-coding system and six months of experience with the system after it was implemented. This information includes the precise time that each trauma patient was admitted to the ED and the time that the patient was discharged to the appropriate hospital department.

The information also includes the severity of the trauma at admission to the ED on a scale of 0 to 75, as well as information related to gender, age, and whether the admission occurred before or after the color-coding system was implemented. The ED director also has information about the color code assigned after the system was initiated that can generally be equated to the severity score assigned at admission to the ED. Trauma scoring 20 or more on the scale would be assigned code red; below 20, code yellow; and not on the scale, no color.

Inferential Statistics, Analysis of Variance, and Regression

The question the ED director wishes to address is how she can use her data to determine whether the color-coding system has reduced the time that trauma victims spend in the ED before being discharged to the appropriate hospital department. At the simplest level, this is a question that can be addressed by using a statistic called the *t* test for comparing two different groups. At a more complex level, the ED director can address the question of whether any difference in waiting time in the ED can be seen as at all related to changes in severity levels of patients before or after the color-coding scheme was introduced. She can also examine whether other changes in the nature of the patients who arrived as trauma victims before and after the introduction of the color-coding scheme might be the cause of possible differences in waiting time. These questions can be addressed by using regression analysis.

Two Caveats of Statistics: Establishing a Significant Difference and Causality

Two caveats regarding the use of statistics apply directly to this example. The first is that no statistical analysis may be needed at all if the waiting time after the initiation of the color-coding scheme is clearly shorter than the waiting time before. Suppose, for example, that the average waiting time before the color-coding scheme was three hours from admission to the ED to discharge to hospital department, and that after the initiation of the scheme, the average waiting time was forty-five minutes. In this scenario, no statistical significance tests would be required to show that the color-coding scheme was associated with a clear improvement in waiting time. Furthermore, it is likely that the color-coding scheme would not only become a permanent part of the ED armamentarium of the university hospital but would be adopted widely by other hospitals as well.

However, suppose that after the initiation of the color-coding scheme the average waiting time in the ED was reduced from three hours to two hours and fifty minutes. A statistical test (probably the *t* test) would show whether 170 minutes of waiting was actually less, statistically, than 180 minutes. Although such a small difference may seem to have little practical significance, it may represent a statistically significant difference. In such a case, the administrator would have to decide whether to retain an intervention (the color-coding scheme) that had a statistical, but not a practical, effect.

The second caveat to the use of statistics is the importance of understanding that a statistical test cannot establish causality. It might be possible, statistically, to show that the color-coding scheme was associated with a statistical reduction in waiting time. But in the absence of a more rigorous study design, it is not possible to say that the color-coding scheme actually *caused* the reduction in waiting time. In a setting such as this, where measurements are taken before and after some intervention (in this case, the color-coding scheme), a number of factors other than the color-coding scheme might have accounted for an improvement in waiting time.

The very recognition of the problem and consequent concern by ED physicians and nurses may have had more effect on waiting time than the color-coding scheme itself. But this is not a question that statistics, per se, can resolve. Such questions may be resolved in whole or in part by the nature of a study design. A double-blind, random clinical trial, for example, is a very powerful design for resolving the question of causality. But, in general, statistical analysis alone cannot determine whether an observed result has occurred *because of* a particular intervention. All that statistical analysis can do is establish whether two events (in this case, the color-coding scheme and the improvement in waiting time) are or are not independent of each other. This notion of independence will come up many more times, and, in many ways, it is the focus of much of this book.

Length of Stay, Readmission Rates, and Cost per Case in a Hospital Alliance

The ever-increasing costs of providing hospital services have sparked a keen interest on the part of hospital administrators in practical mechanisms that can account