First Things First

USING **STATISTICS**
"The Price of Admission"

It's the year 1900, and you are a promoter of theatrical productions, in the business of selling seats for individual performances. Using your knowledge and experience, you establish a selling price for the performances, a price you hope represents a good trade-off between maximizing revenues and not driving away demand for your seats. You print up tickets and flyers, place advertisements in local media, and see what happens. After the event, you review your results and consider the effects of your choices on those results.

Tickets sold very quickly? Next time perhaps you can charge more. The event failed to sell out? Perhaps next time you could charge less or take out more advertisements to drive demand. If you lived over 100 years ago, that's about all you could do.

Jump ahead about 85 years. You're using computer systems that enable you to sell more categories of tickets, such as tickets for premium-priced seat locations. As customers buy tickets over the phone, you can monitor sales through schedule summary reports and, perhaps, add or subtract performance dates using the information in those reports.

Jump ahead to today. Your fully online ticketing system updates seat inventory and enables you to use **dynamic pricing** that automatically alters seat prices based on factors such as increased demand. You also now have the flexibility to set new pricing tables or add special categories that are associated with times of peak demand. Through your sales system you have gained insights about your customers, such as where they live, what other tickets they buy, and their appropriate demographic traits. Because you know more about your customers, you can make your advertising and publicity more efficient by aiming your messages at the types of people more likely to buy your tickets. By using social media networks and other online media, you can also learn almost immediately who is noticing and responding to your advertising messages. You might even run experiments online presenting your advertising in two different ways and seeing which way sells better.

Your current self has capabilities that allow you to be a more effective promoter than any older version of yourself. But just how much better? Turn the page.

CONTENTS

"The Price of Admission"

FTF.1 Think Differently About Statistics

FTF.2 Business Analytics: The Changing Face of Statistics

FTF.3 Starting Point for Learning Statistics

FTF.4 Starting Point for Using Software

FTF.5 Starting Point for Using Microsoft Excel

EXCEL GUIDE

TABLEAU GUIDE

OBJECTIVES

- Statistics is a way of thinking that can lead to better decision making
- Statistics requires analytical skills and is an important part of your business education
- Recent developments such as the use of business analytics and "big data" have made knowing statistics even more critical
- The DCOVA framework guides your application of statistics
- The opportunity business analytics represents for business students

Now Appearing on Broadway ... *and* Everywhere Else

In early 2014, Disney Theatrical Productions surprised Broadway when reports revealed that Disney's *17*-year-old *The Lion King* had been the top-grossing Broadway show in 2013. How could such a long-running show earn so much while being so old? Broadway producers "knew" that grosses for a show decline over time and by year 12 (2009) weekly grosses for *The Lion King* had dropped about 25%. But four years *later,* grosses were up 67%, and weekly grosses typically exceeded even the grosses of the opening weeks of the show, adjusted for inflation!

While heavier advertising and some ticket pricing changes helped, the major reason for this change was something else: combining business acumen with an informed application of *business statistics and analytics* to better solve the problem of selling tickets. As a producer of the newest musical at the time said, "We make educated predictions on price. Disney, on the other hand, has turned this into a science" (Healey).

Disney had followed the plan of action that this book presents. It had collected its daily and weekly results and summarized them, using techniques this book introduces in the next three chapters. Disney then analyzed those results by performing experiments and tests on the data collected, using techniques that later chapters introduce. In turn, insights from the results of those analyses led to developing a new interactive seating map that allowed customers to buy tickets for specific seats and permitted Disney to adjust the pricing of each seat for each performance. The whole system was constantly reviewed and refined, using the semiautomated methods that Chapter 17 introduces. The end result was a ticket-selling method that outperformed previous methods used.

student TIP

From other business courses, you may recognize that Disney's system uses dynamic pricing.

Five years after *The Lion King* surprised Broadway, the show continues to average about the same grosses it did in 2014 and still has many weeks in which it grosses $2 million or more, once a rarely achieved Broadway benchmark. However, the show is no longer the top grossing show because another show, about a partially obscure Secretary of the Treasury has grosses that dwarf *The Lion King*'s respectable sales. The producers of *Hamilton* have fully applied the techniques that Disney first used and created a show whose weekly grosses average more than *$3* million and, even after running for four years, can manage record-setting weekly grosses of more than $4 million! As a *Wall Street Journal* article has noted, "It is boom time on Broadway" (Passy).

FTF.1 Think Differently About Statistics

The "Using Statistics" scenario suggests, and the Disney example illustrates, that modern-day information technology has allowed businesses to apply statistics in ways that could not be done years ago. This scenario and example reflect how this book teaches you about statistics. In these first two pages, you may notice

- the lack of calculation details and "math."
- the emphasis on enhancing business methods and management decision making.
- that none of this seems like the content of a middle school or high school statistics class you may have taken.

You may have had some prior knowledge or instruction in *mathematical statistics*. This book discusses *business statistics*. While the boundary between the two can be blurry, business statistics emphasizes business problem solving and shows a preference for using software to perform calculations.

One similarity that you might notice between these first two pages and any prior instruction is *data*. **Data** are the facts about the world that one seeks to study and explore. Some data are unsummarized, such as the facts about a single ticket-selling transaction, whereas other facts, such as weekly ticket grosses, are **summarized**, derived from a set of unsummarized data. While you may think of data as being numbers, such as the cost of a ticket or the percentage that weekly grosses have increased in a year, do not overlook that data can be non-numerical as well, such as ticket-buyer's name, seat location, or method of payment.

Statistics: A Way of Thinking

Statistics are the methods that allow you to work with data effectively. Business statistics focuses on interpreting the results of applying those methods. You interpret those results to help you enhance business processes and make better decisions. Specifically, business statistics provides you with a formal basis to summarize and visualize business data, reach conclusions about that data, make reliable predictions about business activities, and improve business processes.

You must apply this way of thinking correctly. Any "bad" things you may have heard about statistics, including the famous quote "there are lies, damned lies, and statistics" made famous by Mark Twain, speak to the errors that people make when either misusing statistical methods or mistaking statistics as a substitution for, and not an enhancement of, a decision-making process. (Disney Theatrical Productions' success was based on *combining* statistics with business acumen, not *replacing* that acumen.)

DCOVA Framework

To minimize errors, you use a framework that organizes the set of tasks that you follow to apply statistics properly. Five tasks comprise the **DCOVA framework**:

- **D**efine the data that you want to study to solve a problem or meet an objective.
- **C**ollect the data from appropriate sources.
- **O**rganize the data collected, by developing tables.
- **V**isualize the data collected, by developing charts.
- **A**nalyze the data collected, reach conclusions, and present the results.

You must always do the **D**efine and **C**ollect tasks before doing the other three. The order of the other three varies, and sometimes all three are done concurrently. In this book, you will learn more about the **D**efine and **C**ollect tasks in Chapter 1 and then be introduced to the **O**rganize and **V**isualize tasks in Chapter 2. Beginning with Chapter 3, you will learn methods that help complete the **A**nalyze task. Throughout this book, you will see specific examples that apply the DCOVA framework to specific business problems and examples.

Analytical Skills More Important Than Arithmetic Skills

The business preference for using software to automate statistical calculations maximizes the importance of having analytical skills while it minimizes the need for arithmetic skills. With software, you perform calculations faster and more accurately than if you did those calculations by hand, minimizing the need for advanced arithmetic skills. However, with software you can *also* generate inappropriate or meaningless results if you have not fully understood a business problem or goal under study or if you use that software without a proper understanding of statistics.

Therefore, using software to create results that help solve business problems or meet business goals is *always* intertwined with using a framework. And using software does not mean memorizing long lists of software commands or how-to operations, but knowing how to review, modify, and possibly create software solutions. If you can analyze what you need to do and have a general sense of what you need, you can always find instructions or illustrative sample solutions to guide you. (This book provides detailed instructions *as well as* sample solutions for every statistical activity discussed in end-of-chapter software guides and through the use of various downloadable files and sample solutions.)

If you were introduced to using software in an application development setting or an introductory information systems class, do not mistake building applications from scratch as being a necessary skill. A "smart" smartphone user knows how to use apps such as Facebook, Instagram, YouTube, Google Maps, and Gmail effectively to communicate or discover and use information and has no idea how to construct a social media network, create a mapping system, or write an email program. Your approach to using the software in this book should be the same as that smart user. Use your analytical skills to focus on being an effective user and to understand *conceptually* what a statistical method or the software that implements that method does.

Statistics: An Important Part of Your Business Education

Until you read these pages, you may have seen a course in business statistics solely as a required course with little relevance to your overall business education. In just two pages, you have learned that statistics is a way of thinking that can help enhance your effectiveness in business—that is, applying statistics correctly is a fundamental, global skill in your business education.

In the current data-driven environment of business, you need the general analytical skills that allow you to work with data and interpret analytical results regardless of the discipline in which you work. No longer is statistics only for accounting, economics, finance, or other disciplines that directly work with numerical data. As the Disney example illustrates, the decisions you make will be increasingly based on data and not on your gut or intuition supported by past experience. Having a well-balanced mix of statistics, modeling, and basic technical skills as well as managerial skills, such as business acumen and problem-solving and communication skills, will best prepare you for the workplace today … *and* tomorrow (Advani).

FTF.2 Business Analytics: The Changing Face of Statistics

Of the recent changes that have made statistics an important part of your business education, the emergence of the set of methods collectively known as business analytics may be the most significant change of all. **Business analytics** combine traditional statistical methods with methods from management science and information systems to form an interdisciplinary tool that supports fact-based decision making. Business analytics include

- statistical methods to analyze and explore data that can uncover previously unknown or unforeseen relationships.
- information systems methods to collect and process data sets of all sizes, including very large data sets that would otherwise be hard to use efficiently.
- management science methods to develop optimization models that support all levels of management, from strategic planning to daily operations.

In the Disney Theatrical Productions example, statistical methods helped determine pricing factors, information systems methods made the interactive seating map and pricing analysis possible, and management science methods helped adjust pricing rules to match Disney's goal of sustaining ticket sales into the future. Other businesses use analytics to send custom mailings to their customers, and businesses such as the travel review site tripadvisor.com use analytics to help optimally price advertising as well as generate information that makes a persuasive case for using that advertising.

Generally, studies have shown that businesses that actively use business analytics and combine that use with data-guided management see increases in productivity, innovation, and competition (Advani). Chapter 17 introduces you to the statistical methods typically used in business analytics and shows how these methods are related to statistical methods that the book discusses in earlier chapters.

"Big Data"

Big data is a collection of data that cannot be easily browsed or analyzed using traditional methods. Big data implies data that are being collected in huge volumes, at very fast rates or velocities (typically in near real time), and in a variety of forms that can differ from the structured forms such as records stored in files or rows of data stored in worksheets that businesses use every day. These attributes of volume, velocity, and variety distinguish big data from a "big" (large) set of data that contains numerous records or rows of similar data (Laney). When combined with business analytics and the basic statistical methods discussed in this book, big data presents opportunities to gain new management insights and extract value from the data resources of a business (IBM).

Unstructured Data Big data may also include **unstructured data**, data that have an irregular pattern and contain values that are not comprehensible without additional automated or manual interpretation. Unstructured data take many forms, such as unstructured text, pictures, videos,

and audio tracks, with unstructured text, such as social media comments, getting the most immediate attention today for its possible use in customer, branding, or marketing analyses. Unstructured data can be adapted for use with a number of methods, such as regression, which this book illustrates with conventional, structured files and worksheets. Unstructured data may require performing data collection and preparation tasks beyond those tasks that Chapter 1 discusses. While describing all such tasks is beyond the scope of this book, Section 17.1 includes an example of the additional interpretation that is necessary when working with unstructured text.

FTF.3 Starting Point for Learning Statistics

Statistics has its own vocabulary and learning the precise meanings, or **operational definitions**, of basic terms provides the basis for understanding the statistical methods that this book discusses. For example, *in statistics*, a **variable** defines a characteristic, or property, of an item or individual that can vary among the occurrences of those items or individuals. For example, for the item "book," variables would include the title and number of chapters, as these facts can vary from book to book. For a given book, these variables have a specific value. For *this* book, the value of the title variable would be "Statistics for Managers Using Microsoft Excel," and "19" would be the value for the number of chapters variable. Note that a statistical variable is not an algebraic variable, which serves as a stand-in to represent one value in an algebraic statement and could never take a non-numerical value such as the title of this book.

Using the definition of variable, data, in its statistical sense, can be defined as the set of values associated with one or more variables. In statistics, each value for a specific variable is a single fact, not a list of facts. For example, what would be the value of the variable author for this book? Without this rule, you might say that the single list "Levine, Szabat, Stephan" is the value. However, applying this rule, one would say that the variable has three separate values: "Levine", "Stephan", and "Szabat". This distinction of using only *single-value data* has the practical benefit of simplifying the task of entering data for software analysis.

Using the definitions of data and variable, the definition of statistics can be restated as the methods that analyze the data of the variables of interest. The methods that primarily help summarize and present data comprise **descriptive statistics**. Methods that use data collected from a small group to reach conclusions about a larger group comprise **inferential statistics**. Chapters 2 and 3 introduce descriptive methods, many of which are applied to support the inferential methods that the rest of the book presents.

Statistic

The previous section uses *statistics* in the sense of a collective noun, a noun that is the name for a collection of things (methods in this case). The word statistics also serves as the plural form of the noun statistic, as in "one uses methods of descriptive statistics (collective noun) to generate descriptive statistics (plural of the singular noun)." In this sense, a **statistic** refers to a value that summarizes the data of a particular variable. For the Disney Theatrical Productions example, the statement "for 2013, weekly grosses were up 67% from 2009" cites a *statistic* that summarizes the variable weekly grosses using the 2013 data—all 52 values.

When someone warns you of a possible unfortunate outcome by saying, "Don't be a statistic!" you can always reply, "I can't be." *You* always represent one value, and a *statistic* always summarizes multiple values. For the statistic "87% of our employees suffer a workplace accident," you, as an employee, will either have suffered or have not suffered a workplace accident. The "have" or "have not" value contributes to the statistic but cannot be the statistic. A statistic can facilitate preliminary decision making. For example, would you immediately accept a position at a company if you learned that 87% of their employees suffered a workplace accident? (Sounds like this might be a dangerous place to work and that further investigation is necessary.)

Can Statistics (*pl.*, statistic) Lie?

The famous quote "lies, damned lies, and statistics" actually refers to the plural form of *statistic* and does not refer to statistics, the field of study. Can any statistic "lie"? No, faulty or invalid

statistics can only be produced through willful misuse of statistics or when DCOVA framework tasks are done incorrectly. For example, many statistical methods are valid only if the data being analyzed have certain properties. To the extent possible, you test the assertion that the data have those properties, which in statistics are called *assumptions*. When an assumption is *violated*, shown to be invalid for the data being analyzed, the methods that require that assumption should not be used.

For the inferential methods that this book discusses in later chapters, you must always look for logical causality. **Logical causality** means that you can plausibly claim something directly causes something else. For example, you wear black shoes today and note that the weather is sunny. The next day, you again wear black shoes and notice that the weather continues to be sunny. The third day, you change to brown shoes and note that the weather is rainy. The fourth day, you wear black shoes again and the weather is again sunny. These four days seem to suggest a strong pattern between your shoe color choice and the type of weather you experience. You begin to think if you wear brown shoes on the fifth day, the weather will be rainy. Then you realize that your shoes cannot plausibly influence weather patterns, that your shoe color choice cannot *logically cause* the weather. What you are seeing is mere coincidence. (On the fifth day, you do wear brown shoes and it happens to rain, but that is just another coincidence.)

You can easily spot the lack of logical causality when trying to correlate shoe color choice with the weather, but in other situations the lack of logical causality may not be so easily seen. Therefore, relying on such correlations by themselves is a fundamental misuse of statistics. When you look for patterns in the data being analyzed, you must *always* be thinking of logical causes. Otherwise, you are misrepresenting your results. Such misrepresentations sometimes cause people to wrongly conclude that all statistics are "lies." Statistics (*pl.*, statistic) are not lies or "damned lies." They play a significant role in *statistics*, the way of thinking that can enhance your decision making and increase your effectiveness in business.

FTF.4 Starting Point for Using Software

learnMORE

About the online supplemental files in Appendix C.

The starting point for using any software is knowledge of basic user interface skills, operations, and vocabulary that Table FTF.1 summarizes and which the supplemental **Basic Computing Skills** online document reviews.

TABLE FTF.1
Basic computing skills

Skill or Operation	Specifics
Identify and use standard window objects	Title bar, minimize/resize/close buttons, scroll bars, mouse pointer, menu bars or ribbons, dialog box, window subdivisions such as areas, panes, or child windows
Identify and use common dialog box items	Command button, list box, drop-down list, edit box, option button, check box, tabs (tabbed panels)
Mouse (or touch) operations	Click, called select in some list or menu contexts and check or clear in some check box contexts; double-click; right-click; drag and drag-and-drop

With such knowledge, learning how to apply any program to a business problem becomes possible. In using Microsoft Excel, this book assumes no more than this starting point. While having prior experience is always useful, readers unfamiliar with Excel can learn about commonly used operations in this book's appendices. Such readers will find instructions for using reusable tools and automated methods, the Excel add-ins Analysis ToolPak and PHStat, throughout the book. Those instructions assume only this starting-point knowledge.

Software-Related Conventions Table FTF.2 summarizes the typographical conventions that this book uses for its software instructions. These conventions provide a concise and clear way of expressing specific basic-user activities and operations.

TABLE FTF.2 Software typographical conventions in this book

Convention	Example
Special key names appear capitalized and in boldface	Press **Enter**. Press **Command** or **Ctrl**.
Key combinations appear in boldface, with key names linked using this symbol: **+**	Enter the formula and press **Ctrl+Enter**. Press **Ctrl+C**.
Menu or Ribbon selections appear in boldface and sequences of consecutive selections are shown using this symbol: **➔**	Select **File ➔ New**. Select **PHStat ➔ Descriptive Statistics ➔ Boxplot**.
Target of mouse operations appear in boldface	Click **OK**. Select **Attendance** and then click the **Y button**.
Entries and the location of where entries are made appear in boldface	Enter **450** in cell **B5**. Add **Temperature** to the **Model Effects** list.
Variable or column names sometimes appear capitalized for emphasis	This file contains the Fund Type, Assets, and Expense Ratio for the growth funds.
Placeholders that express a general case appear in italics and may also appear in boldface as part of a function definition	**AVERAGE (*cell range of variable*)** Replace *cell range of variable* with the cell range that contains the Asset variable.
Names of data files mentioned in sections or problems appear in a special font but appear in boldface in end-of-chapter Guide instructions	Retirement Funds Open the **Retirement Funds workbook**.
When current versions of Excel differ in their user interface, alternate instructions appear in a second color immediately following the primary instructions	In the Select Data Source display, click **Edit** that appears under **Horizontal (Category) Axis Labels**. In Excel for Mac, in the Select Data Source display, click the icon inside the **Horizontal (Categories) axis labels**.

Using Software Properly

Learning to use software *properly* can be hard as software has limited ways to provide feedback for user actions that are invalid operations. In addition, no software will ever know if its users are following the proper procedures for use. Exhibit FTF.1 presents a list of guiding principles that, ideally, would govern readers' usage of any software, not just the software used with this book. These principles will minimize the chance of making errors and lessen the frustration that often occurs when these principles are overlooked.

EXHIBIT FTF.1

Principles of Using Software Properly

Ensure that software is properly updated. Users who manage their own computers often overlook the importance of ensuring that all installed software is up to date.

Understand the basic operational tasks. Take the time to master the tasks of starting the application, loading and entering data, and how to select or choose commands in a general way.

Understand the statistical concepts that an application uses. Not understanding those concepts can lead to making wrong choices in the application and can make interpreting results difficult.

Know how to review software use for errors. Review and verify that the proper data preparation procedures (see Chapter 1) have been applied to the data before analysis.

(*continued*)

EXHIBIT FTF.1 (*continued*)

Verify that the correct procedures, commands, and software options have been selected. For information entered for labeling purposes, verify that no typographical errors exist.

Seek reuse of preexisting solutions to solve new problems. Build solutions from scratch only as necessary, particularly if using Excel in which errors can be most easily made. Some solutions, and almost all Excel solutions that this book presents, exist as models or templates that can *and should* be reused because such reuse models best practice.

Understand how to organize and present information from the results that the software produces. Think about the best ways to arrange and label the data. Consider ways to enhance or reorganize results that will facilitate communication with others.

Use self-identifying names, especially for the files that you create and save. Naming files Document 1, Document 2, and so on will impede the later retrieval and use of those files.

FTF.5 Starting Point for Using Microsoft Excel

Microsoft Excel is the data analysis component of Microsoft Office that evolved from earlier electronic spreadsheets first used for financial accounting applications. As such, Excel can be an inadequate tool for *complex* data analysis, even as the program is a convenient way to visually examine data details as well as learn foundational applications of business statistics.

Excel uses **worksheets** to display the contents of a data set and as a means to enter or edit data. Worksheets are containers that present tabular arrangements of data, in which the intersections of rows and columns form **cells**, boxes into which individual entries are made. In the simplest use of Excel, cell entries are single data values that can be either text or numbers. When entering data for a data set to be analyzed, data for each variable are placed in their own column, a vertical group of cells. By convention, the initial row of such columns stores the variable name associated with the column data. Columns so prepared are called either variables, variable columns, or just columns in the context of using Microsoft Excel.

studentTIP

The Pearson Excel add-in **PHStat** automates the use of the Excel Guide workbooks, further simplifying the operational details of using Excel for data analysis. Appendix H explains more about PHStat and how it can be used with this book.

Worksheet cells can also contain **formulas**, instructions to process data or to compute cell values. Formulas can include **functions** that simplify certain arithmetic tasks or provide access to advanced processing or statistical features. Formulas play an important role in designing **templates**, *reusable solutions* that have been previously tested and verified. A hallmark feature of this book is the inclusion of **Excel Guide workbooks**, most of which exist as templates, and all of which simplify the operational details of using Excel for data analysis while demonstrating the application of formulas and functions for such analysis. Appendix C includes a list of the Excel Guide workbooks that this book uses.

Excel stores worksheet data and results as one file called a **workbook**. Workbooks can contain multiple worksheets and chart sheets, sheets that display visualizations of data. Because workbooks contain collections, best practice places data and the results of analyzing that data on separate worksheets, a practice that the Excel Guide workbooks and the Excel data workbooks for this book reflect. Therefore, instructions for using Excel that appear in end-of-chapter Excel Guides often begin with "Open to the ***specific* worksheet** in the ***specific* workbook**" as in "Open to the **DATA worksheet** of the **Retirement Funds workbook**."

Appendix C contains a list of Excel data workbooks, many of which are single-worksheet workbooks.

An "Excel file" is *always* a workbook file, even if it only contains one worksheet. Excel workbook files have either the file extension **.xlsx**, .**xlsm**, or .**xlam**. The file extension tells users whether the workbook contains data only (.xlsx) or data plus macro or add-in instructions (the other two), which may require an additional step to open, as Appendix D explains.

Tableau Differences Readers supplementing Excel with Tableau Public need to understand that while Tableau also uses workbooks to store one or more worksheets, Tableau defines these things differently. A Tableau workbook contains worksheets that present tabular and visual summaries, each of which is associated with a *data source*. A **data source** is a pointer to data that can be a complex collection of data or be as simple as an Excel worksheet (the data sources that this book uses for examples). A Tableau workbook can contain more than one data source, and each data source can be used to create multiple visual or tabular summaries.

Tableau workbooks can also include dashboards, a concept that Chapter 17 discusses.

Data sources can be viewed and column formulas can be used to define new columns, but individual values *cannot be edited* because Tableau Public workbooks do not store the data themselves. Therefore, if using a Tableau workbook, the associated data source must be available if data need to be edited or used for additional analyses.

More About the Excel Guide Workbooks

The Excel Guide workbooks contain reusable templates and model solutions for specific statistical analyses. All workbooks are organized similarly, with the data to be analyzed placed on its own page. When working with templates, readers only enter or paste in their data and never enter or edit formulas, thereby greatly reducing the chance that the worksheet will produce erroneous results. When working with a model worksheet solution, readers enter or paste in their data and edit or copy certain formulas, an additional step that the Pearson Excel add-in PHStat can automate.

Not starting from scratch minimizes the chance of errors, and using templates or, when necessary, model solutions reflects best business practice. Allowing individuals to create new solutions from scratch in business can create risk and led to internal control violations. For example, in the aftermath of the 2012 "London Whale" trading debacle, which caused an estimated loss of at least $6.2 billion, JP Morgan Chase discovered a worksheet that could greatly miscalculate the volatility of a trading portfolio because a formula divided a value by the sum of some numbers and not the average of those numbers (Ewok, Hurtado).

Reusability Templates and model solutions are reusable because they are capable of recalculation. In worksheet **recalculation**, results displayed by formulas automatically change as the data to which the formulas refer change. Reusability also minimizes internal control risk by allowing worksheet formulas and macro to be first tested and audited for correctness before being used for information and decision-making purposes.

Excel Skills That Readers Need

To use Excel effectively with this book, readers will need to know how to make and edit cell entries, how to open to a particular worksheet in a workbook, how to print a worksheet, and how to open and save files. Readers without these skills should review the introduction to these skills that starts in the Excel Guide for this chapter and continues in Appendix B. Readers not using PHStat will also need to modify model worksheet solutions, especially for inferential topics such as ANOVA and regression that later chapters of this book discuss.

Excel Guide Instructions Chapter Excel Guides provide separate instructions for working with the Excel Guide workbooks directly or for using PHStat for specific statistical analyses that chapter sections discuss. Instructions for working with the workbooks directly have the subhead ***Workbook*** and instructions for using PHStat have the subhead ***PHStat***. Topics that do not include the *Workbook* subhead require solutions that would be too difficult to modify for all but advanced users of Excel to perform. Note that for advanced inferential topics, *Workbook* instructions may contain the advisory that the instructions are for intermediate or advanced Excel users. Such instructions are not designed for readers who are casual or novice Excel users.

studentTIP

Readers of past editions of this book who chose not to use PHStat often found the ToolPak instructions, where they exist, more convenient to use than the *Workbook* instructions for a topic.

For selected topics, Excel Guide instructions include the third subhead ***Analysis ToolPak*** that provides instructions for using the Excel Analysis Toolpak add-in. Because they explore chapter examples or solve end-of-section or end-of-chapter problems, readers can use a mix of *Workbook*, *PHStat*, or *Analysis ToolPak* instructions because the instructions for a topic have been designed to create identical results.

REFERENCES

Advani, D. "Preparing Students for the Jobs of the Future." *University Business* (2011). **bit.ly/1gNLTJm**.

Davenport, T., J. Harris, and R. Morison. *Analytics at Work*. Boston: Harvard Business School Press, 2010.

Healy, P. "Ticker Pricing Puts 'Lion King' atop Broadway's Circle of Life." *New York Times, New York edition*, March 17, 2014, p. A1. **nyti.ms.1zDkzki**.

Hurtado, P. "The London Whale." *Bloomberg QuickTake*, February 23, 2016. **bloom.bg/2qnsXKM**.

Ewok, J. "The Importance of Excel." *The Baseline Scenario*, February 9, 2013, **bit.ly/1LPeQUy**.

IBM Corporation. "What Is Big Data?" **www.ibm.com/big-data/us/en/**.

Laney, D. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Stamford, CT: META Group. February 6, 2001.

Levine, D., and D. Stephan. "Teaching Introductory Business Statistics Using the DCOVA Framework." *Decision Sciences Journal of Innovative Education* 9 (Sept. 2011): 393–398.

Passy, C. "Broadway Ticket Sales Sizzle This Summer." *The Wall Street Journal,* August, 20, 2018. **on.wsj.com/2X3ZWRP**.

KEY TERMS

big data 34
cells 38
data 32
business analytics 34
DCOVA framework 33
descriptive statistics 35
formula 38
function 38
inferential statistics 35
logical causality 36
operational definition 35
recalculation 39
reusability 39
statistic 35
statistics 33
summarized data 32
template 38
unstructured data 34
variable 35
workbook 38
worksheets 38

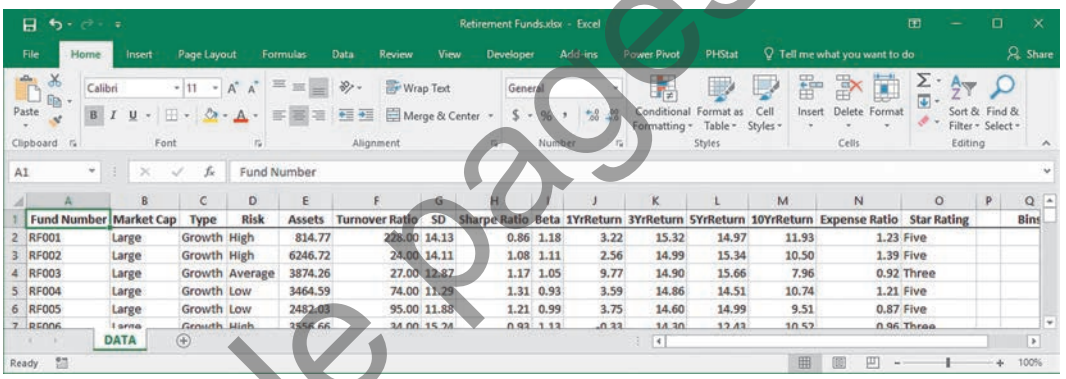
EG.1 GETTING STARTED with EXCEL

Opening Excel displays a window that contains the Office Ribbon tabs above a worksheet area. When a workbook is opened, Excel displays the name of the workbook centered in the title bar. The top of the worksheet area contains a formula bar that enables one to edit the contents of the currently selected cell (cell A1 in the illustration). At the bottom of the worksheet grid, Excel displays a sheet tab that identifies the current worksheet name (DATA). In workbooks with more than one sheet, clicking a sheet tab makes the worksheet named by the tab the current worksheet.

The illustration below shows the Retirement Funds workbook, one of the Excel data workbooks for this book.

student TIP

Excel sometimes displays a task pane in the worksheet area that presents formatting and similar choices.



Fund Number	Market Cap	Type	Risk	Assets	Turnover Ratio	SD	Sharpe Ratio	Beta	1YrReturn	3YrReturn	5YrReturn	10YrReturn	Expense Ratio	Star Rating
RF001	Large	Growth	High	814.77	228.00	14.13	0.86	1.18	3.22	15.32	14.97	11.93	1.23	Five
RF002	Large	Growth	High	6246.72	24.00	14.11	1.08	1.11	2.56	14.99	15.34	10.50	1.39	Five
RF003	Large	Growth	Average	3874.26	27.00	12.87	1.17	1.05	9.77	14.90	15.66	7.96	0.92	Three
RF004	Large	Growth	Low	3464.59	74.00	11.29	1.31	0.93	3.59	14.86	14.51	10.74	1.21	Five
RF005	Large	Growth	Low	2482.03	95.00	11.88	1.21	0.99	3.75	14.60	14.99	9.51	0.87	Five

EG.2 ENTERING DATA

In Excel, enter data into worksheet columns, starting with the leftmost, first column, using the cells in row 1 to enter variable names. Avoid skipping rows or columns as such skipping can disrupt or alter the way certain Excel procedures work. Complete a cell entry by pressing **Tab** or **Enter**, or, if using the formula bar to make a cell entry, by clicking the **check mark icon** in the formula bar. To enter or edit data in a specific cell, either use the cursor keys to move the cell pointer to the cell or select the cell directly.

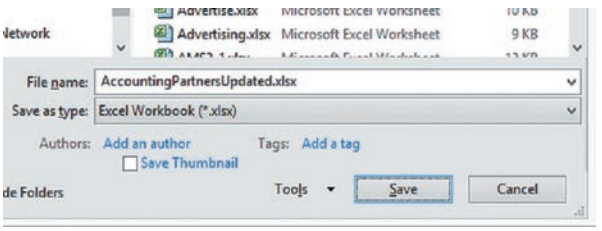
When using numbers as row 1 variable headings, precede the number with an apostrophe. Pay attention to special instructions in this book that note specific orderings of variable columns that are necessary for some Excel operations. When in doubt, use the DATA worksheets of the Excel Guide Workbooks as the guide for entering and arranging variable data.

EG.3 OPEN or SAVE a WORKBOOK

Use **File ➔ Open** or **File ➔ Save As**.

Open and **Save As** use similar means to open or save the workbook by name while specifying the physical device or network location and folder for that workbook. Save As dialog boxes enable one to save a file in alternate formats for programs that cannot open Excel workbooks (**.xlsx** files) directly. Alternate formats include a simple text file with values delimited with tab characters, **Text (Tab delimited) (*.txt)** that saves the contents of the current worksheet as a simple text file, and **CSV (Comma delimited) (*.csv)** that saves worksheet cell values as text values that are delimited with commas. Excels for Mac list these choices as **Tab Delimited Text (.txt)** and **Windows Comma Separated (.csv)**.

The illustration below shows the part of the Save As dialog box that contains the **Save as type** drop-down list. (Open dialog boxes have a similar drop-down list.) In all Windows Excel versions, you can also select a file format in the Open dialog box. Selecting **All Files (*.*)** from the drop-down list can list files that had been previously saved in unexpected formats.



To open a new workbook, select **File ➔ New** (**New Workbook** in Excel for Mac). Excel displays a new workbook with one or more blank worksheets.

EG.4 WORKING with a WORKBOOK

Use **Insert** (or **Insert Sheet**), **Delete**, or **Move or Copy**.

Alter the contents of a workbook by adding a worksheet or by deleting, copying, or rearranging the worksheets and chart sheets that the workbook contains. To perform one of these operations, right-click a sheet tab and select the appropriate choice from the shortcut menu that appears.

To add a worksheet, select **Insert**. In Microsoft Windows Excel, you also click **Worksheet** and then click **OK** in the Insert dialog box. To delete a worksheet or chart sheet, right-click the sheet tab of the worksheet to be deleted and select **Delete**.

To copy or rearrange the position of a worksheet or chart sheet, right-click the sheet tab of the sheet and select **Move or Copy**. In the Move or Copy dialog box, first select the workbook and the position in the workbook for the sheet. If copying a sheet, also check **Create a copy**. Then click **OK**.

EG.5 PRINT a WORKSHEET

Use **File ➔ Print**.

Excel prints worksheets and chart sheets, not workbooks. When you select **Print**, Excel displays a preview of the currently opened sheet in a dialog box or pane that enables worksheet selections to be made. Adjust the print formatting of the worksheet(s) to be printed by clicking **Page Setup**. Typically, in the Page Setup dialog box, one might click the **Sheet** tab and then check or clear the **Gridlines** and **Row and column headings** checkboxes to include or delete these headings from a printout.

EG.6 REVIEWING WORKSHEETS

When reviewing worksheets remember that what is displayed in cells may be the result of either the recalculation of formulas or cell formatting. A cell that displays 4 might contain the value 4, might contain a formula calculation that results in the value 4, or might contain a value such as 3.987 that has been formatted to display as the nearest whole number.

To display and review all formulas, press **Ctrl+`** (grave accent). Excel displays the *formula view* of the worksheet, revealing all formulas. (Pressing **Ctrl+`** a second time restores the worksheet to its normal display.) Note that the Excel Guide Workbooks contain one or more FORMULAS worksheets that provide a second way of viewing all formulas.

The template and model solutions for this book format specific cells by changing text attributes or the background color of cells, with cells containing numeric results typically formatted to display four decimal places. Appendix B discusses these as well as other common worksheet and chart sheet formatting operations.

EG.7 IF YOU USE the *WORKBOOK* INSTRUCTIONS

Excel Guide *Workbook* instructions use the word *display* as in the "Format Axis display" to refer to a user interaction that may be presented by Excel in a **task pane** or a **two-panel dialog box** ("Format Axis task pane" or the "Format Axis dialog box"). Task panes open to the side of the worksheet and can remain onscreen indefinitely. Initially, some parts of a pane may be hidden and an icon or label must be clicked to reveal that hidden part to complete a *Workbook* instruction. Two-panel dialog boxes that open over the worksheet must be *closed* to continue to use Excel. The left panel of such dialog boxes are always visible and clicking entries in the left panel makes visible one of the right panels. (Click the system close button at the top right of a task pane or dialog box to close the display and remove it from the screen.)

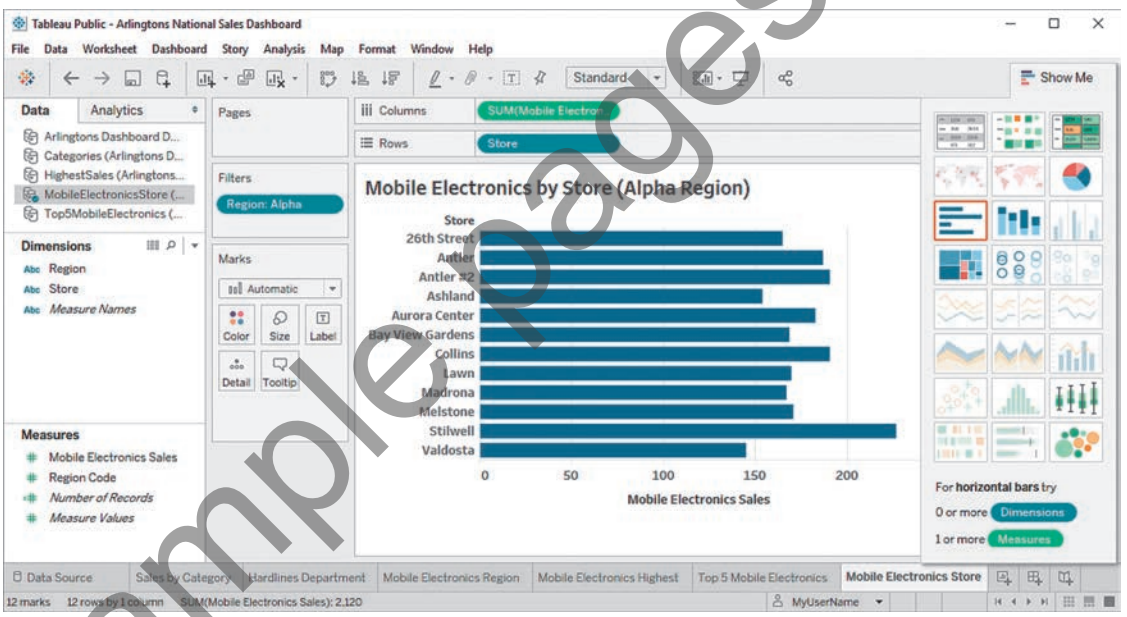
Current Excel versions can vary in their menu sequences. Excel Guide instructions show these variations as parenthetical phrases. For example, the menu sequence, "select **Design** (or **Chart Design**) ➔ **Add Chart Element**" tells you to first select **Design** *or* **Chart Design** to begin the sequence and then to continue by selecting **Add Chart Element**. (Microsoft Windows Excel uses **Design** and Excel for Mac uses **Chart Design**.)

For the current Excel versions that this book supports (see the FAQs in Appendix G), the *Workbook* Instructions are generally identical. Occasionally, individual instructions may differ significantly for one (or more) versions. In such cases, the instructions that apply for multiple versions appear first, in normal text, and the instructions for the unique version immediately follows in this text color.

TG.1 GETTING STARTED with TABLEAU

The Tableau Guides for this book feature Tableau Public, version 2019, also known as the Tableau Desktop Public Edition. Tableau Public uses a drag-and-drop interface that will be most familiar to users of the Excel PivotTable feature, which Chapter 2 discusses.

Tableau Public uses workbooks to store tabular and visual worksheet summaries, dashboards, and stories. Opening Tableau Public displays the main Tableau window in which the contents of a workbook can be viewed and edited. The main window shown below displays the Mobile Electronics Store worksheet of the Arlingtons National Sales Dashboard Tableau workbook that Chapter 17 uses as an example.



The main Tableau window contains a menu bar and toolbar, a tabbed left area that presents data and formatting details, several special areas that Section TG.4 explains, the worksheet display area, and the Show Me gallery that displays tabular and visual summaries appropriate for the data in the worksheet area. Worksheet tabs appear under these areas, along with a tab for the current data source to the left of the tabs and icon shortcuts for a new worksheet, new dashboard, and new story, respectively, to the right of the tabs. The bottom of the window displays status information at the left, the current signed-in user ("MyUserName"), a drop-down button to sign out of Tableau Public, and four media controls that permit browsing through the worksheet tabs of the current workbook.

The Data tab shown above displays the workbook data sources (five) and lists the "Dimensions" and "Measures," the variable columns of the data source for the worksheet, summary measures, and calculated values for the MobileElectronicsStore data source associated with the displayed worksheet. (Section TG1.1 explains more about the significance of dimensions and measures.)

For the worksheet shown above, the variable Store was dragged-and-dropped in the Rows *shelf* and the variable Mobile Electronics Sales was dragged-and-dropped in the Column *shelf*. The worksheet also contains a *filter* for the Region column that selects only those rows in the source worksheet in which the value in the Region column is Alpha. (To create the bar chart, **horizontal bars icon** was selected from the Show Me gallery.)

TG.2 ENTERING DATA

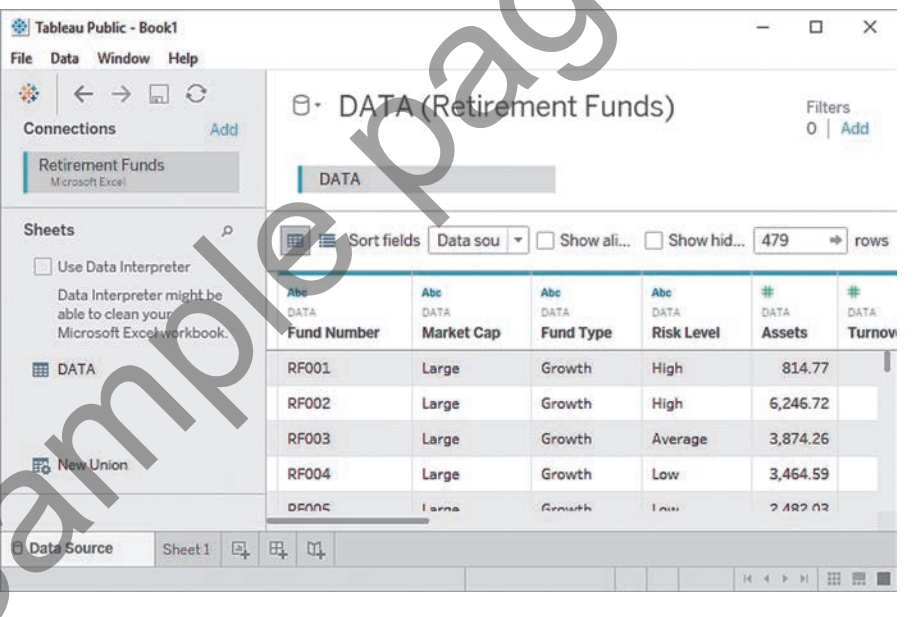
As Section FTF.5 explains, one defines data sources, which are pointers to Excel worksheets and other data files. Tableau does not support the direct entry of data values.

TG.3 OPEN or SAVE a WORKBOOK

Use **File → Open** or **File → Open from Tableau Public**.
Use **File → Save to Tableau Public As**.

Use Open to import simple data sources such as Microsoft Excel workbooks or text files or to open a *Tableau* workbook that has been previously downloaded and saved. Most Tableau Guide instructions begin by opening an Excel workbook to avoid the limitations of Tableau Public that Appendix Section G.5 discusses.

To open a new workbook, select **File → New**. For a new workbook, the Data tab will display the hyperlink **Connect to Data**. Clicking the hyperlink displays the Connect panel from which data sources can be retrieved. When opening an Excel workbook that contains more than one data worksheet, each Excel worksheet being used must be defined as a separate data source. This Connect panel also appears when opening an Excel workbook directly using the file open command. The illustration below shows the data source linked to the Data worksheet of the Excel Retirement Funds workbook.



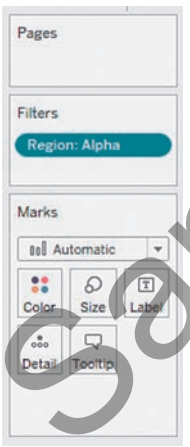
Using the Open from Tableau Public or Save to Tableau Public As command requires a valid Tableau online account. (Accounts are complimentary but require registration.) Using these commands means signing into a Tableau Public account and retrieving (or storing) a Tableau workbook from that account or from an account that has been shared. Save to Tableau Public As stores the Tableau workbook in the account and opens a web browser to display the workbook and to permit its downloading to a local computing device. In paid-subscription editions of Tableau Desktop, these open and save commands appear in the Server menu and not in the File menu.

Although not used in the Tableau Guides, Tableau Public permits *join* and *union* operations to combine columns from two data sources. **Join** operations combine two tables, typically by matching values in a variable column that both original tables share. **Union** operations add rows. Union operations require that tables share columns that hold values for the same variables. Joins and unions can solve problems that arise from seeking to perform an analysis of data on a set of variables stored in two different places, such as two different Excel data worksheets in the same Excel workbook.

TG.4 WORKING with DATA

Tableau Desktop uses the term **data field** to refer to what this book calls a variable or, in some contexts, a column. What Excel calls a formula function, Tableau calls an aggregation. What Excel calls a formula, Tableau calls an aggregate calculation. Tableau also invents its own vocabulary for several user interface elements in the worksheet window that readers may know under more common names. Knowing this vocabulary can be helpful when consulting the Tableau help system or other references.

Tableau calls the Pages, Filters, Columns, and Rows areas, all shown in the Section TG.1 illustration, *shelves*. The Marks area seems to be a shelf, but for reasons that would be only self-evident to a regular user of Tableau, the Marks area is called a *card*. **Shelves** are places into which things can be placed or dropped, such as the *pills* that have been placed in the Filters, Row, and Columns shelves. A **pill** represents some data and is so named because it reminds some of a medicinal capsule. In its simplest form, a pill represents a data field in the data source. However, pills can represent a filtering operation, such as the Filters shelf pill in the illustration below, or a calculated result, similar to a worksheet or data table formula. Pills can be either blue or green, reflecting the type of numerical data, discrete or continuous (see Section TG1.1), or red, reflecting an error condition.



In the Section TG.1 illustration, the Store Dimension has been dragged to the Rows shelf, creating the Store pill and the Mobile Electronics Sales Measure has been dropped on the Columns shelf, creating the SUM(Mobile Electronics Sales) pill (slightly truncated, see figure detail below). Dropping a measure name on a shelf creates an aggregation (formula function). The SUM aggregation pill sums mobile electronic sales values by rows (each store). In the special case where each store is represented by only one row, there is no actual summing of values.

TG.5 PRINT a WORKBOOK

Tableau Public does not contain a print function that would print worksheets. (Commercial versions of Tableau do.) To print a tabular or visual summary, use a screen capture utility to capture the display for later printing. For online worksheets displayed in a web browser, use the print function of the browser.