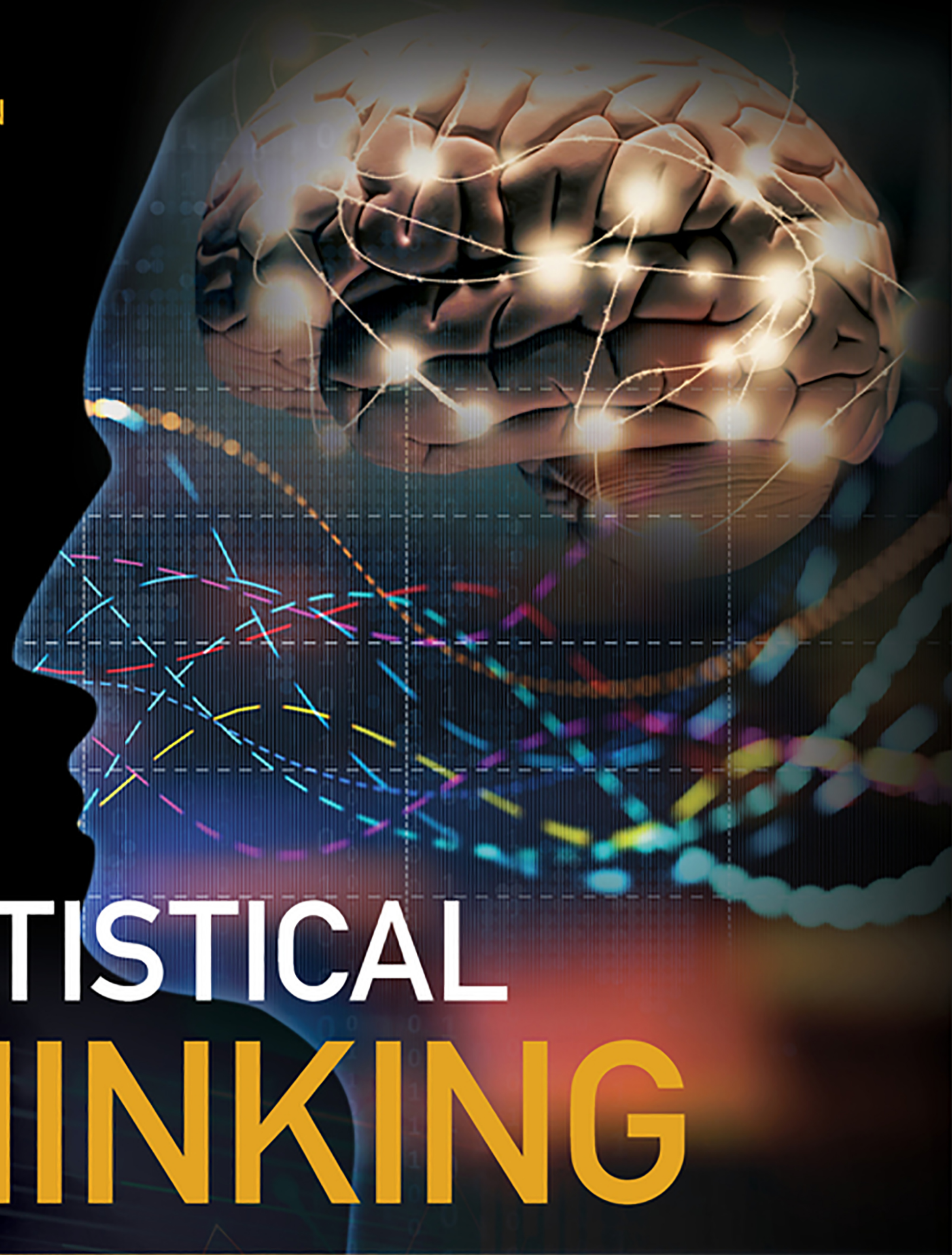
THIRD EDITION



STATISTICAL THINKING

IMPROVING BUSINESS PERFORMANCE

ROGER W. HOERL AND RONALD D. SNEE

WILEY

Statistical Thinking

Wiley & SAS Business Series

The Wiley & SAS Business Series presents books that help senior-level managers with their critical management decisions.
Titles in the Wiley and SAS Business Series include:

Activity-Based Management for Financial Institutions: Driving Bottom-Line Results by Brent Bahnub

Branded! How Retailers Engage Consumers with Social Media and Mobility by Bernie Brennan and Lori Schafer

Business Analytics for Customer Intelligence by Gert Laursen

Business Analytics for Managers: Taking Business Intelligence beyond Reporting by Gert Laursen and Jesper Thorlund

Business Intelligence Competency Centers: A Team Approach to Maximizing Competitive Advantage by Gloria J. Miller, Dagmar Brautigam, and Stefanie Gerlach

Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy by Olivia Parr Rud

Case Studies in Performance Management: A Guide from the Experts by Tony C. Adkins

CIO Best Practices: Enabling Strategic Value with Information Technology, Second Edition by Joe Stenzel

Credit Risk Assessment: The New Lending System for Borrowers, Lenders, and Investors by Clark Abrahams and Mingyuan Zhang

Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring by Naeem Siddiqi

Customer Data Integration: Reaching a Single Version of the Truth, by Jill Dyche and Evan Levy

Demand-Driven Forecasting: A Structured Approach to Forecasting by Charles Chase

Enterprise Risk Management: A Methodology for Achieving Strategic Objectives by Gregory Monahan

Executive's Guide to Solvency II by David Buckham, Jason Wahl, and Stuart Rose

Fair Lending Compliance: Intelligence and Implications for Credit Risk Management by Clark R. Abrahams and Mingyuan Zhang

Foreign Currency Financial Reporting from Euros to Yen to Yuan: A Guide to Fundamental Concepts and Practical Applications by Robert Rowan

Information Revolution: Using the Information Evolution Model to Grow Your Business by Jim Davis, Gloria J. Miller, and Allan Russell

Manufacturing Best Practices: Optimizing Productivity and Product Quality by Bobby Hull.

Marketing Automation: Practical Steps to More Effective Direct Marketing by Jeff LeSueur

Mastering Organizational Knowledge Flow: How to Make Knowledge Sharing Work by Frank Leistner

Performance Management: Finding the Missing Pieces (to Close the Intelligence Gap) by Gary Cokins

Performance Management: Integrating Strategy Execution, Methodologies, Risk, and Analytics by Gary Cokins

Retail Analytics: The Secret Weapon by Emmett Cox

Social Network Analysis in Telecommunications by Carlos Andre Reis Pinheiro

The Business Forecasting Deal: Exposing Bad Practices and Providing Practical Solutions by Michael Gilliland

The Data Asset: How Smart Companies Govern Their Data for Business Success by Tony Fisher

The Executive's Guide to Enterprise Social Media Strategy: How Social Networks Are Radically Transforming Your Business by David Thomas and Mike Barlow

The New Know: Innovation Powered by Analytics by Thornton May

The Value of Business Analytics: Identifying the Path to Profitability by Evan Stubbs

Visual Six Sigma: Making Data Analysis Lean by Ian Cox, Marie A. Gaudard, Philip J. Ramsey, Mia L. Stephens, and Leo Wright

For more information on any of the above titles, please visit www.wiley.com.

Statistical Thinking

Improving Business Performance

Third Edition

Roger W. Hoerl and Ronald D. Snee

WILEY

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

Edition History
Duxbury Press (1e, 2002); John Wiley and Sons, Inc. (2e, 2012).

Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data

Names: Hoerl, Roger Wesley, 1957- author. | Snee, Ronald D., author.
Title: Statistical thinking : improving business performance / Roger W. Hoerl, Ronald D. Snee.
Description: Third edition. | Hoboken, New Jersey : John Wiley & Sons, Inc., [2020] | Series: Wiley SAS business series | Includes bibliographical references and index.
Identifiers: LCCN 2020013168 (print) | LCCN 2020013169 (ebook) | ISBN 9781119605713 (hardback) | ISBN 9781119605720 (adobe pdf) | ISBN 9781119605737 (epub)
Subjects: LCSH: Commercial statistics.
Classification: LCC HF1012 .H64 2020 (print) | LCC HF1012 (ebook) | DDC 658.4/033—dc23
LC record available at https://lccn.loc.gov/2020013168
LC ebook record available at https://lccn.loc.gov/2020013169

Cover Design: Wiley
Cover Images: active human brain © PM Images/ Getty Images, Artificial Intelligence concept © MF3d/Getty Images, financial stock market graph © whiteMocca/Shutterstock

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

*To the memory of Arthur E. Hoerl, Horace P. Andrews,
and Ellis R. Ott—great teachers from whom we
learned much about the theory and use
of statistical thinking*

Table of Contents

Preface

This book utilizes a state-of-the-art approach to teaching introductory business statistics, which has been tested for more than eight years in the classroom. We believe it fills an important void, one that many have been concerned about for some time. This preface describes the philosophy and content of this book, and explains the uniqueness of the third edition.

It is generally accepted that in the latter half of the twentieth century, introductory statistics was often poorly taught, being reduced to a collection of formulas divorced from real data, and especially from real problems. Fortunately, a number of forward-thinking statisticians, scientists, and engineers saw the limitations in this approach, and set out to do better. A number of enhancements have been gradually introduced since then and continue to this day. For example, consider the Curriculum Guidelines for Undergraduate Programs in Statistical Science (https://www.amstat.org/asa/files/pdfs/EDU-guidelines2014-11-15.pdf), or the Guidelines for Assessment and Instruction in Statistics Education Report (https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx).

These more modern approaches provided important enhancements to the formulaic approach, including:

More utilization of real data

Balance between theory and practice

A mix of instructional media, beyond traditional lectures

Broad use of technology in the classroom, including statistical software

Incorporating effective communication of the results of statistical studies.

We certainly agree with, and also use, these modern approaches to introductory statistical education. However, we feel that more progress is needed in order for students to grasp the full power and relevance of statistics. Therefore, in our first and second editions we incorporated the key points noted above, and also emphasized the following principles, which in our opinion are critical but often overlooked, even in modern texts:

- Discussion of the "big picture," what we are actually trying to accomplish in statistical applications, prior to discussing the methods themselves. Understanding the "why" before the "what" and "how" helps students understand how statistics can help them in their own endeavors, beyond the statistics class.

- Acknowledging that activity and work in all fields can and should be viewed as a process that can be studied and improved. The process provides the critical context for any data, so that it can be properly understood and analyzed. This view leads naturally to utilization of statistical methods for process improvements and problem-solving that produces real impact. Chapter 4 on "Understanding Business Processes" is found in no other statistics texts of which we are aware.
- Emphasizing the dynamic nature of real processes, and the practical inability to obtain the "random samples" so often assumed in texts. Random sampling is an ideal that is rarely achieved in practice.
- Using case studies and applications to introduce and justify theory, rather than presenting theory first in isolation, and then presenting numerical illustrations of the theory. Our experience and cognitive theory both suggest that most students learn best going from the tangible to the abstract, rather than vice versa.
- Utilization of real business examples and problems, including incorporation of real business data sets encountered in the authors' practices.
- Illustration of the sequential nature of statistical applications, rather than the "one-shot studies" so often presented in textbooks.
- Focusing on the tools most frequently used in practice.

Our primary motivation for a third edition was to publish what we feel is the most modern introductory statistics text on the market, and most reflective of current statistical practice. By this, we mean that it addresses the major issues and topics that statisticians are talking about at conferences, applying on the job, and writing about in journals. Secondly, we feel that this edition has broken new ground by incorporating a separate chapter dedicated to data and the need to understand data pedigree (data quality). The topic of data pedigree is only starting to be widely understood within the profession, but with the growth of data science and Big Data sets, it will grow more significant as an issue the profession has to deal with. Thirdly, we also feel we have the most modern approach to think about variation, one that incorporates outliers as well as random and systematic variation. We feel this is a more holistic approach to variation, and one that we have been teaching to students for a while, with very good results.

The specific updates and improvements incorporated in the third edition are as follows.

Proving a state-of-the-art introductory business statistics text, incorporating the major issues the profession is currently focused on. These include:

The relationship between statistical thinking and data science

The importance of understanding the data pedigree (data quality)

The need to understand data within the context of the process that produced it, including the measurement system involved

The importance of data visualization, integrated with numerical approaches, and now running through the text as a theme

The proper role and interpretation of *p*-values

The distinction between practical and statistical significance

The emerging discipline of statistical engineering, which is now being discussed in both quality and statistical journals

The proper integration of subject matter theory with data-based approaches and models.

Including a new separate chapter on data, data pedigree (quality), tools for acquiring data, and analysis of the measurement system as part of evaluating data pedigree. We are not aware of any other textbook on the market with a chapter specifically focused on data and data pedigree.

Providing a framework for variation that includes outliers as well as systematic (explainable) and random variation. That is, outliers, which occur regularly in virtually all application areas, are not ignored, as in some texts, or viewed as "oops, something went wrong," as in others. Rather, in this broader definition of outliers, they fit within the overall model for variation introduced in Chapter 1 and are referred to in each subsequent chapter.

Several chapters have been completely rewritten, based on several years of experience teaching with the first and second editions, including student input. This includes updating various graphs and tables.

We trust that readers will find the third edition to be an example of the application of statistical thinking to improve a textbook.

We would also like to thank and acknowledge our colleagues from Wiley who helped make this 3rd edition possible, including Rahini Devi, Karen Weller, Purvi Patel, and numerous others.

Roger W. Hoerl
Niskayuna, NY

Ronald D. Snee
Newark DE

Introduction to JMP*

JMP is desktop data analysis software from SAS, the world's leading provider of analytics solutions for industry. JMP is easy to learn and use and contains a very broad collection of tools for data analysis and visualization. It also works well with data in other formats, including Microsoft Excel, and is available for both Windows and Macintosh operating systems. A free 30-day trial that you can easily download and install to use for the examples in this book is available at www.jmp.com/trial.

In this section we will introduce you to some of the essential functions of JMP, including basic navigation, how to import data, how to run basic analyses, and where to get help. You will find additional resources at www.jmp.com/learn and many excellent books at www.jmp.com/books.

WHY JMP?

In one package, JMP contains all the basic graphing and analysis tools found in spreadsheets as well as more advanced platforms for regression, design of experiments, and quality and predictive analytics. JMP is designed around the workflow of the data analyst and provides several important advantages to the user. The first of these is that JMP guides you to the appropriate analysis for your data. The results are always driven by the type of data you have and the general purpose of your analysis. JMP then provides contextual options, allowing you to dive deeper into the analysis.

The second advantage is that graphs nearly always accompany statistical results; the graphs are presented first, followed by the numerical results. Note that JMP also provides a separate Graph menu that contains additional visualization tools that are independent of numerical results. Another important advantage is that graphs in every platform are dynamically linked to the data, allowing one to explore relationships visually and to perform data management tasks on the fly. We are confident that JMP will save you time and yield better results.

*Some of the material in the Introduction is adapted from Curt Hinrichs and Chuck Boiler, *JMP® Essentials: An Illustrated Step-by-Step Guide for New Uses* (Cary, NC: SAS Institute, 2010).

JMP MENUS

At the top of the JMP window, you see a series of menus (File, Edit, Tables, etc.). These menus are used to open or import data, to edit or restructure data, to design an experiment and to create graphs and analyses. There is also a valuable source for assistance through the Help menu, which is discussed later. Note that while we are illustrating JMP on the Windows platform, Macintosh instructions are nearly identical. (See Figure A.)

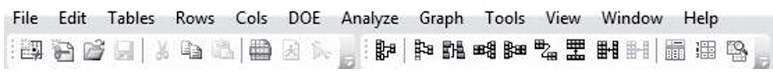


FIGURE A JMP Menu Bar

The menus are organized in a logical sequence from left to right:

- **File** is where you go to open or import data and to save, print, or exit JMP. It is also where you can customize the appearance or settings within JMP through Preferences.
- **Edit** appears on the Mac home window and, on Windows, in individual data tables and reports (but not on the Windows Home window). Edit provides the usual cut, clear, copy, paste, and select functions, as well as undo, redo and special JMP functions.
- **Tables** provide the tools to manage, summarize, and structure your data.
- **DOE** contains the Design of Experiments tools and the sample size and power calculators.
- **Analyze** contains the analysis tools that generate both graphs and statistics. It serves as the home for all of JMP's statistical tools from the most basic to advanced.
- **Graph** contains graphical tools that are independent of statistics (at least initially). Graphs in this menu include basic charts to advanced multivariable and animated visualization tools.
- **Tools** allows you to transform your mouse into a help tool, a selection tool, a brushing tool or scrolling tool, and much more.
- **View** lets you view or hide windows or toolbars.
- **Window** helps you manage windows within JMP.
- **Help** provides resources for learning and using JMP. The Help menu provides access to the learning resources (including all of the documentation) that you will use as you expand your knowledge of JMP and its features and learn about statistics.

IMPORTING DATA

Importing data is similar to opening any file from a desktop application. In Windows, click **File > Open** to launch the Open dialog window. Near the bottom of the window you will notice a file type button that allows (see Figure B) you to select from a variety of data formats that JMP can read natively. If you know the format of your data, select that format to see available files of that type.

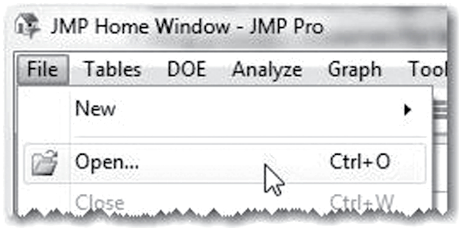


FIGURE B File > Open

Select or highlight the file and click **Open** (see Figure C).

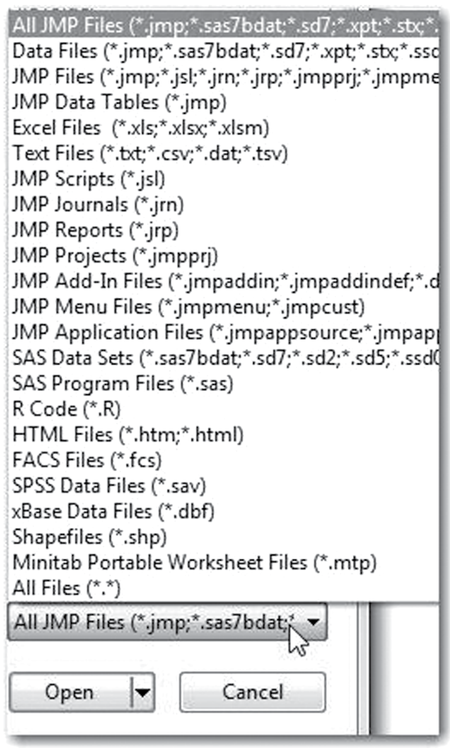


FIGURE C JMP Import Formats

JMP can also import data extracted from databases via ODBC. For more information about these and other data importing functions, click **Help > JMP Help>Using JMP**.

THE JMP DATA TABLE

The JMP Data table is similar to any spreadsheet with a few important differences. JMP requires your data to be structured in a standard form, where variables are in columns and observations are in rows. Whether you are importing data from another source or creating a new data table, make sure this format is in place.

The data table also contains metadata or information about your data. The most important of these is the modeling type of your variables, which is displayed in the middle (Columns) panel on the left-hand side of the data table. The modeling type will drive the type of results you get from an analysis, meaning that JMP only produces statistics and graphs that are suitable for the type of data you are working with and the analysis at hand. You can change the modeling type to another appropriate alternative by simply clicking on the icon and selecting the desired modeling type. (See Figure D.)

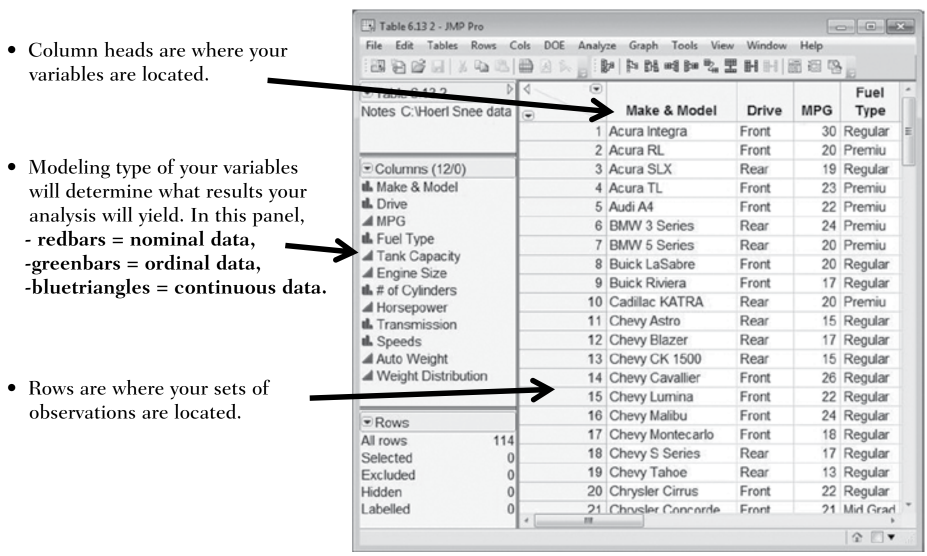


FIGURE D JMP Data Table

THE ANALYZE MENU

As noted earlier, the Analyze menu is where you will find the statistical tools in JMP. Nearly all of the statistical results you generate in this menu will also generate an associated graph, and that graph will appear first. The menu is designed to support the objective of your analysis and provides a very logical sequence to the order in which the items appear. The most basic and general tools are at the top of the menu, and as you move down the menu, the tools become more advanced or specific.

JMP contains few menu items relative to its capabilities because the combination of your modeling type and analysis objective will always narrow down and produce the appropriate graphs and statistical results. Let us take a look at some of the items on the Analyze menu. In the top section, you find the following (see Figure E):

- **Distribution** (for univariate statistics). A good starting point with any data set. See what each column looks like and generate summary statistics. Confidence intervals and hypothesis tests for one variable (Chapter 9).
- **Fit Y by X** (for bivariate statistics). Explore relationships between any two variables (one Y and one X). Simple regression, one-way, contingency, and so forth (Chapter 7).
- **Fit Model.** A robust platform for multiple regression and general modeling (more than one Y or X) (Chapter 7).

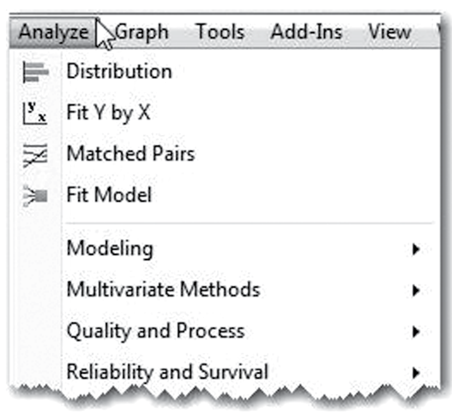


FIGURE E JMP Analyze Menu

The next items (beginning with Modeling) are families of tools that contain submenus with more specific functions.

The **Modeling** menu contains platforms for data mining (Partition and Neural), time series forecasting, and categorical data analysis among others (see Figure F). The **Multivariate Methods** menu contains common multivariate tools, such as Clustering, Factor Analysis and Correlations. While these two menu items are beyond the scope of this book, the interested reader can find more information at **Help > JMP Help>Multivariate Methods**.

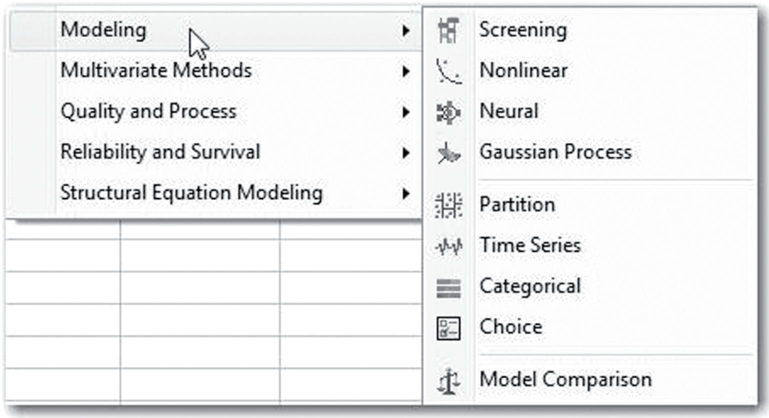


FIGURE F JMP Modeling Menu

The **Quality and Process** menu was recently added and has consolidated many of the quality-related tools in a logical manner. Control Chart Builder allows you to create control charts in a drag-and-drop manner and will be illustrated later (Figure G). More information is available at **Help > JMP Help > Quality and Process Methods**.

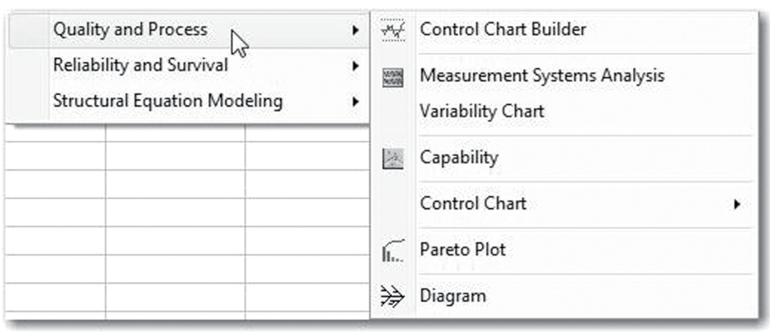


FIGURE G JMP Quality and Process Menu

JMP DIALOG WINDOWS

When you select most Analyze menu items, a dialog window will appear consisting of three main components (Figure H):

1. On the left side of the window, the **Select Columns** area contains the variables in your data table that you can select for your analysis.
2. In the middle section, the **Roles** in which you'd like to cast those variables (you can do this by dragging them into the role, or by selecting them and then clicking the button for that role).
3. On the right side are a series of **Actions** you can take. Click **OK** to launch the specified analysis.

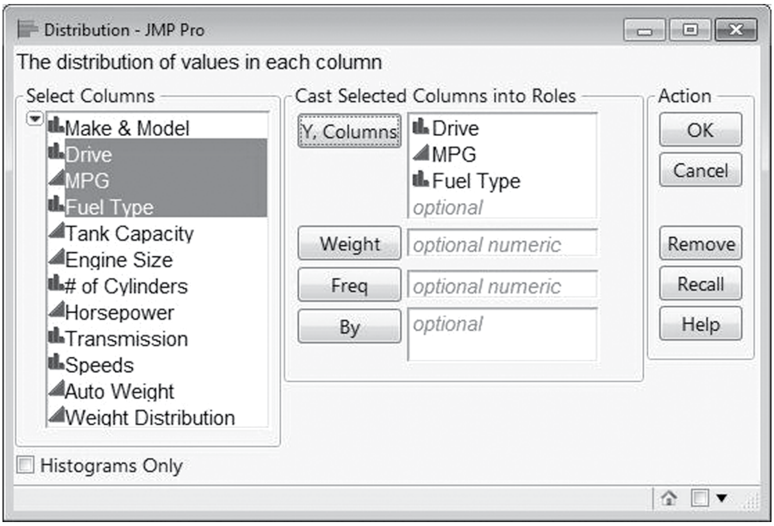


FIGURE H JMP Dialog Window

THE GRAPH MENU

The graph menu contains a wide variety of data visualization platforms. Unlike the analyze menu where you generate both statistical results and graphs, the Graph menu generates only graphs of your data or models (at least initially) (Figure I).

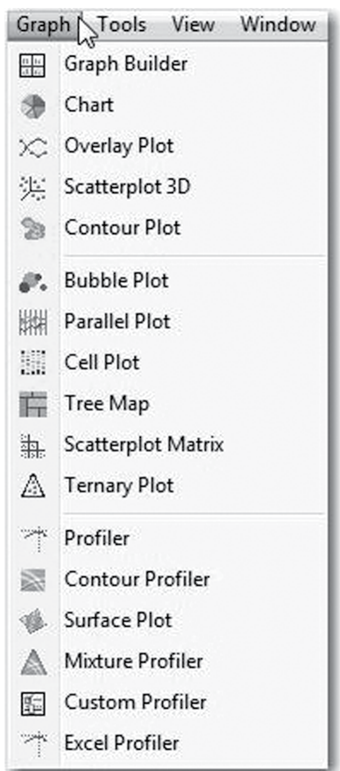


FIGURE I JMP Graph Menu

One very powerful platform is the first item on the menu, Graph Builder. If you are ever looking for that "best" graph of your data, Graph Builder provides an easy and unique way to find it. Graph Builder provides a canvas with surrounding zones in which to drag-and-drop your variables (Figure J). We will show you this later.

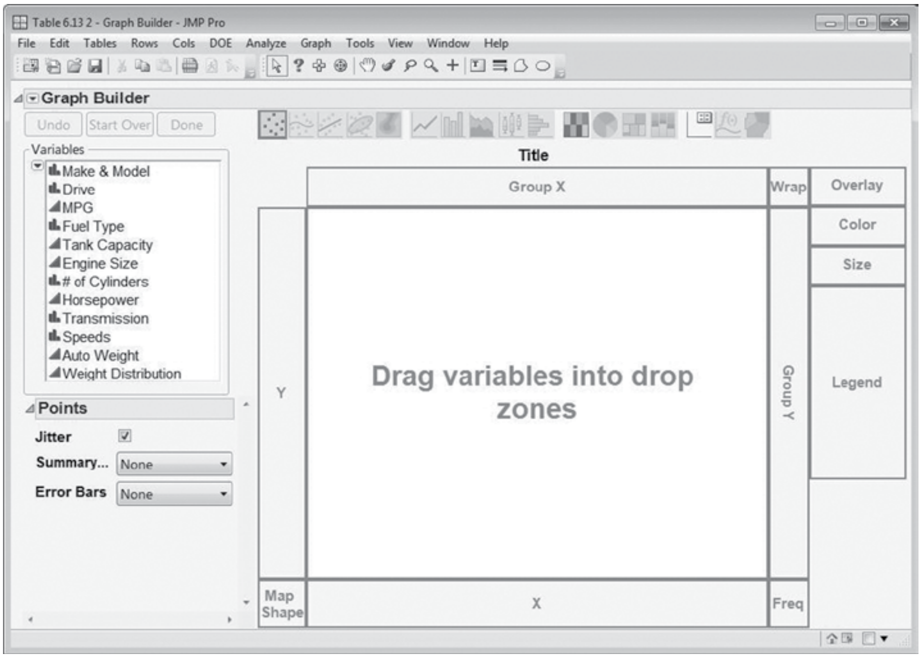


FIGURE J JMP Graph Builder

THE DOE MENU

The DOE (Design of Experiments) dialog box (Chapter 8) is accessed through the DOE menu, which provides a wide range of design options. If you are running experiments, this is where you will begin. The platform contains the Custom Designer, which is recommended for new users of DOE because it helps the user find the most efficient design for the types of responses and factors in the experiment (Figure K).

Since experiments are designed before you collect data, the DOE dialog windows are designed to help you define the response(s) and factors in the experiment (Figure L). Once defined, JMP will create the design that will guide your data collection.

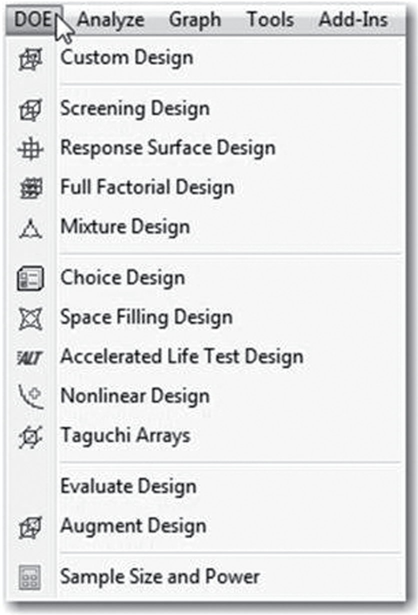


FIGURE K JMP DOE Menu

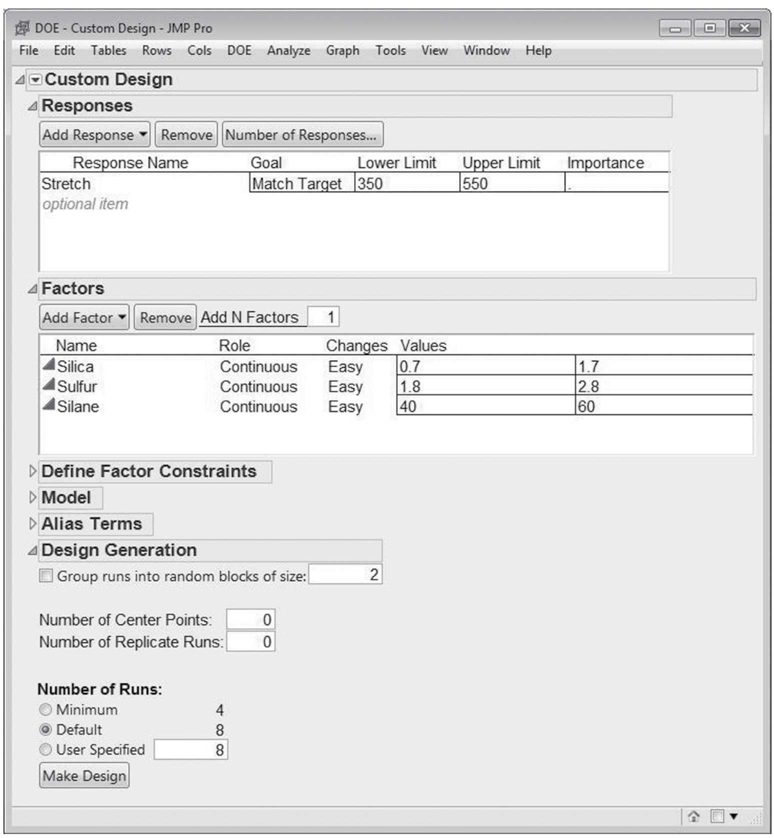


FIGURE L JMP DOE Dialog Window

THE TOOLS MENU

There are several tools in JMP that transform your mouse into a powerful and versatile tool. With the Tools menu, you can easily move results or graphs into other documents (e.g. PowerPoint, Word, the Web, etc.), annotate your output or understand what the results mean. When one of these options is selected, you will notice that your cursor will transform into the icon associated with the item. Let us summarize what the really important ones do (Figure M).

- **Arrow**. This is a general-purpose selection tool and the default setting used 99% of the time.
- **Help**. Select this and move the cursor to whatever you have a question about and click. This will take you to the documentation for the item in question.
- **Selection**. This will allow you to copy and paste a graph or results into another program or document. Select the item, right-click, select Copy and Paste into the desired document (or just drag the selected image into the desired document).

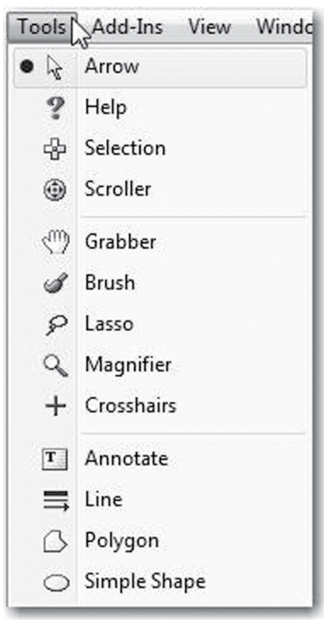


FIGURE M JMP Tools Menu

USING JMP

Now that you have been introduced to the essentials of JMP, let us run through a few examples. For more detailed step-by-step instructions for specific statistical methods, consult the one-page guides at www.jmp.com/learn.

1. Finding Confidence Intervals for One Mean (Chapter 9):

a. Open the data file **Table E7.5**

b. Click **Analyze > Distribution**.

c. Highlight or drag **MPG** to **Y, Columns**.

The 95% Confidence Interval for the Mean will display by default under Summary Statistics. For a hypothesis test for the mean, or additional options, click the red triangle for the variable (circled in Figure N).

2. Creating a graph using Graph builder (Figure O):

a. Open the data file **Table E7.5**

b. Click **Graph > Graph Builder**.

c. Drag **MPG** to **Y** and release, and drag **Horsepower** to **X** and release.

d. Drag another variable to Group X or Group Y (here we've put **Transmission** in **Group X** at the top).

e. Click an icon in the pallet to change the type of graph.

f. Click **Done** to produce the graph.

Use Graph Builder to produce histograms, box plots, scatterplots, maps, summary statistics, and more.

3. Running a simple regression model with Fit Y by X (Figure P)

a. Open the data file for **Table E7.5**

b. Click **Analyze > Fit Y by X**.

c. Select **MPG** and click **Y, Response**.

d. Select **Horsepower** and click **X, Factor**.

e. Click **OK**.

f. Under the red triangle select **Fit Line**.

The formula for the line is displayed under Linear Fit. Additional options, such as correlations (select Density Ellipse), are available under the top red triangle. Residuals and other options related to the model are available under the red triangle next to Linear Fit.

To build a regression model with more than one X variable use **Analyze > Fit Model**.

4. Creating a control chart with Control Chart Builder (Figure Q)

   a. Open the data file for **Table 7.4.**

   b. Click **Analyze > Quality and Process > Control Chart Builder**.

   c. Drag **MPG** to **Y** and release.

   d. Drag **Month** to the bottom of the graph pane, the **Subgroup** zone, and release.

   e. Click **Done** to produce the graph.

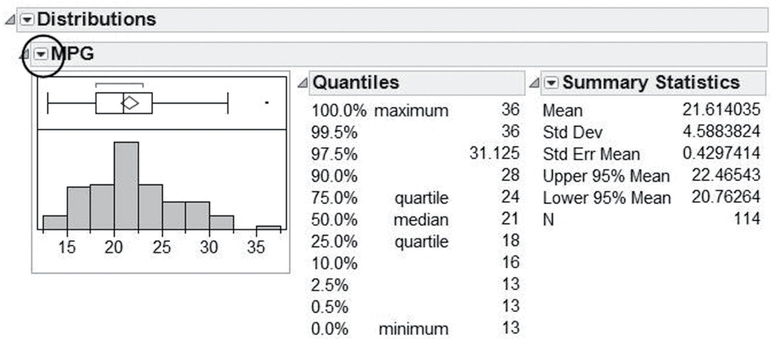


FIGURE N JMP Distribution Output with Circle

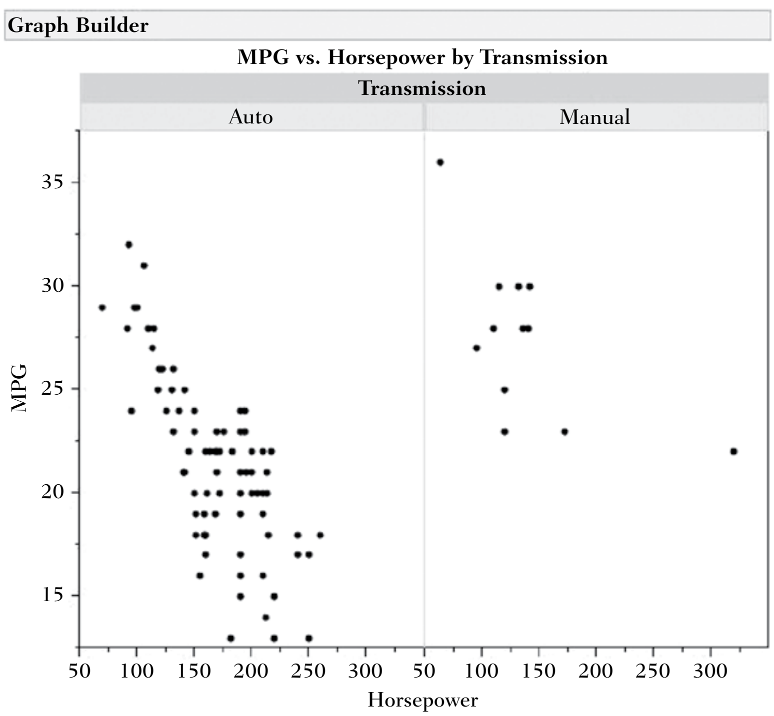
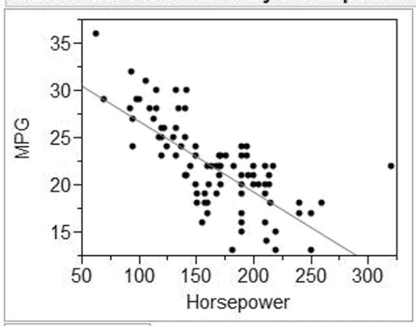


FIGURE O JMP Graph Builder Output

Bivariate Fit of MPG By Horsepower



Linear Fit

Linear Fit

MPG = 34.208637 - 0.0748623*Horsepower

Summary of Fit

Lack Of Fit

Analysis of Variance

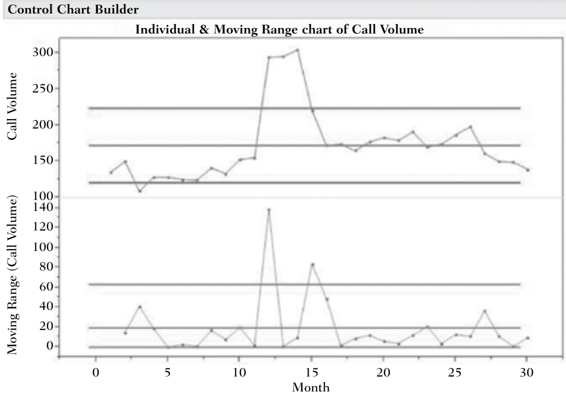
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1266.7140	1266.71	127.5479
Error	112	1112.3036	9.93	**Prob > F**
C. Total	113	2379.0175		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	34.208637	1.153585	29.65	<.0001*
Horsepower	-0.074862	0.006629	-11.29	<.0001*

FIGURE P JMP Fit Y by X Output

Control Chart Builder



Call Volume Limit Summaries

Points plotted	LCL	Avg		Limits Sigma
Individual	120.6433	172.1667	223.6901	Moving Range
Moving Range	0	19.37931	63.30314	Moving Range

FIGURE Q JMP Control Chart Builder Output

If data are subgrouped, an X-Bar chart will be produced. To compute limits for different phases (machines, operators, etc.), drag and drop the variable in the Phase zone at the top.

CHAPTER 1

Need for Business Improvement

If you don't keep doing it better—your competition will.

—Anonymous

In today's global marketplace, success—even survival—hinges on an organization's ability to improve everything it does. In this chapter, we demonstrate why corporations need to improve how they run their businesses and how the use of statistical thinking can improve business performance. Statistical thinking can be applied to both business operations and methods of management.

The main objective of Chapter 1 is to better understand the effect of global competition and technology change on business and other organizations in our society and how this impact is forcing us to improve. You will become familiar with the various approaches to improvement and how statistical thinking plays a role in each of these methods. This will enable you to see how the broad use of statistical thinking can help businesses and other organizations improve.

We begin with a short case study. Generalizing from the case study, we then discuss today's business realities, the need to improve, and the recognition that improving how we work is now part of the job. The need to improve while we accomplish our work is illustrated with an overall model for business improvement. We then briefly review some new management approaches. Common themes that run through these approaches are identified, and the role of statistical thinking in these themes, and hence in the improvement effort, is noted.

TODAY'S BUSINESS REALITIES AND THE NEED TO IMPROVE

Consider the following business scenario. A large publication corporation, Kowalski and Sons, is having trouble with their monthly billing process. They have discovered that it takes about 17 days to send bills out to customers. But there is a lot of variation from billing cycle to billing cycle, with some bills taking much longer than 17 days. Management's expectation is that the billing should be done in less than 10 days with minimal variation. This target is important from both the company's and the customers' point of view. A shorter cycle time for the bills would improve the company's cash flow, and it would allow customers to enter the billing information in their accounting systems promptly so they can close their monthly books sooner. The current situation results in numerous "late" payments, for which Kowalski and their customers often blame each other. Customers complain that other publishers are not as tardy in sending out bills.

Does this sound like a bad situation? Actually, this is a typical situation in many businesses. In fact, when one of the authors consulted on this problem and began to dig deeper, the situation became worse! Assessing the process revealed that three different departments were involved in billing. Each department worked separately, and no one understood the process from beginning to end. When problems occurred, there was a lot of finger pointing: "The problem is not with us, it's with them. If they would clean up their act, the billing process would be okay." Similarly, there were no standard operating procedures—that is, formal, agreed-upon methods of doing the job. Everybody did it their own way. This resulted in a lot of "fire fighting" to keep the bills going out—heroic efforts requiring long hours and shifting priorities.

The one clear advantage was that a quantitative measure to monitor performance did exist: the number of days required to send bills out. Without a clear measure of success, it is difficult—if not impossible—to effectively manage and improve a process.

Traditional business leaders faced with this situation might attempt to assign blame so the persons responsible could be reprimanded. The approach we recommend is just the opposite. Here is how we approached this problem: A systems map was created for the overall process, along with a flowchart of the critical process steps. The systems map identified the responsible departments and the information or materials that flowed back and forth between the groups. The flowchart was used to construct a production schedule for the monthly billing cycle. This schedule showed what had to be done each month by each group along with a timetable for doing so.

Next, critical subprocesses were identified and cycle time measurements were monitored for each of these critical subprocesses as well as for the overall process. These measurements highlighted key problem areas. Cross-functional teams were formed to troubleshoot the process daily and to review the billing process at the end of the cycle. These teams identified problems and suggested procedures for creating and implementing solutions.

Efforts were also made to document the process and the procedures used in its operation. This documentation helped reduce variation in the process and was central to training new employees. A process owner was also assigned. The process owner's job was to care for the "health" of the process by seeing that the various aspects of the process management system were used and improved to handle the changing conditions the process would experience.

Use of this statistical thinking approach significantly improved the billing process. Over a 5-month period, the monthly billing cycle time was reduced from an average of 17 days to about 9.5 days, with less variation. This resulted in annual savings of more than $2.5 million, more satisfied customers, and a less stressful work environment for employees.

The use of statistics in business has grown over the years as a result of political, social, technological, and economic forces that have affected our world economy. Each new force has created a new need for statistics that typically results in new concepts, methods, tools, and applications. For example, World War II created the need for statistical quality control: Munitions needed to be manufactured consistently to very tight tolerances. The need for statistical design of experiments resulted from the demand for major increases in farm production in the early 1900s, which required experimentation with new farming techniques. The application of statistical design of experiments to agricultural productivity has been so effective that world hunger is today a problem of logistics, rather than productivity.

This statistical movement was accelerated both by the former Soviet Union's launch of the *Sputnik* satellite, resulting in a major scientific initiative by the United States, and by the increasing focus on research and development in the chemical and process industries during the 1950s and 1960s.

The U.S. Food, Drug, and Cosmetics Act and the U.S. Environmental Protection Act resulted in increased use of statistics in the pharmaceutical industry and in environmental studies in the 1970s. The advent of the computer also made statistical calculations easier and available to a broader range of people. The 1980s brought a new economic force—global competition—which has created the need to make major changes in how we run our businesses. The need for change is driven by increasing customer demands for more responsive companies and for higher quality products and services at lower costs.

Global competition is affecting economies around the world; indeed we now have a global economy. Evidence of the effects of the global marketplace on the U.S. economy can be seen in the balance of trade and median hourly wages (adjusted for inflation) shown in Figures 1.1 and 1.2. These plots indicate a robust U.S. economy in the 1950s and 1960s, but things clearly changed

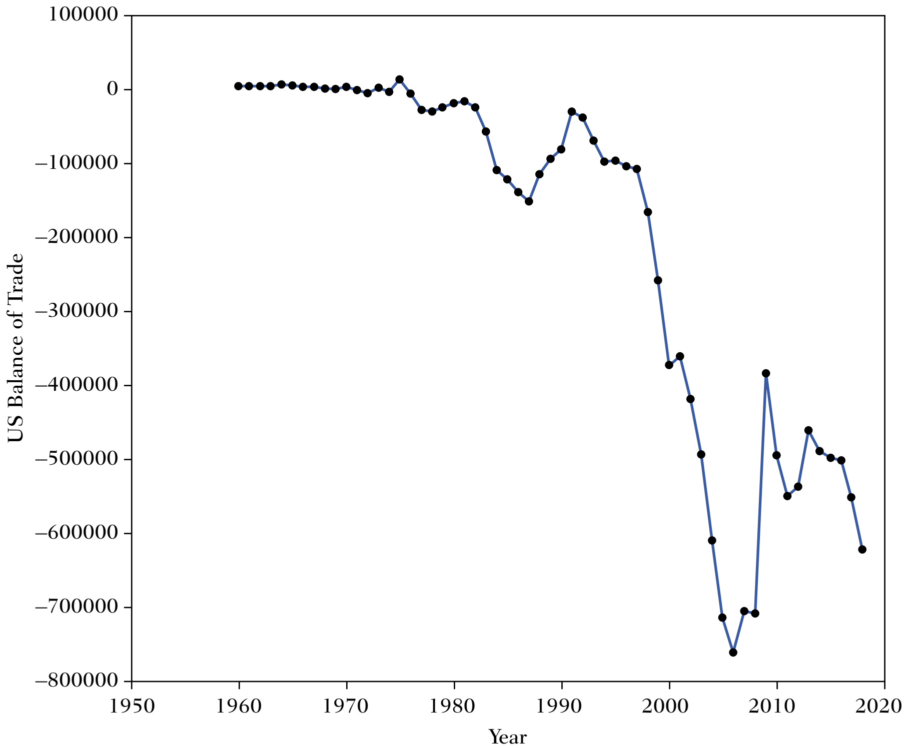


FIGURE 1.1 U.S. Balance of Trade, 1960–2018

in the 1970s and 1980s. Global competition became a serious challenge to the U.S. economy. Figure 1.1 shows that the trade balance of goods and services (exports minus imports) was positive until 1971, when it turned negative. Despite some positive upturns, it has remained significantly negative from the 1980s through the twenty-first century.

In Figure 1.2, we see that the U.S. median hourly earnings adjusted for inflation increased until 1973 and then decreased until the mid-1990s, when it began to increase again. But as of 2018, U.S. median hourly earnings have only reached the levels of the early 1970s. This indicates a declining standard of living for the United States as a whole over most of this time period.

Global competition has had an impact on the U.S. economy in other ways as well. Companies find it difficult to compete, which results in layoffs, downsizing, mergers, and bankruptcies. Many of 1960s Fortune 500 companies are not in business today. For many years, the General Electric Company (GE) was the only surviving member of the original set of 30 leading industrial corporations that made up the Dow Jones Industrial Average, founded in 1896. When the Dow dropped GE in 2018 due to poor business performance, none of the original 30 top industrials remained. In the new millennium, the Internet

FIGURE 1.2 U.S. Median Hourly Earnings, 1964–2018 (2018 Dollars)

and social media are driving forces. Many "traditional" businesses are being replaced by digital competitors with a radically different business model, such as Amazon replacing "brick and mortar" stores, Kindle and other electronic readers replacing physical books, or Netflix and Amazon Prime replacing movie theaters. Does anyone still buy paper maps in the era of global positioning systems (GPS)?

The changes in industry and Internet technologies have combined to usher in "Industry 4.0," or the fourth phase of the industrial revolution. The first phase was the industrial revolution in the early nineteenth century, enabled by water and steam power. The second phase was brought on by the assembly line and electrification in the early twentieth century. The third phase was due to computerization and automation in the late twentieth century.

The fourth phase, occurring now, has been brought about by massive data sets ("Big Data"), the direct connection of equipment through the Internet (the "Internet of Things"), and cyber-physical systems, which integrate humans, software, and physical devices. Think of Uber or Lyft: these cyber-physical systems only work because there is a physical device—a car, humans making decisions on buying or selling rides, and of course the software that allows these people and devices to communicate and track one another in real time.

The changes taking place in the global economy have ripple effects throughout society, including government, education, health care, and non-profit organizations. For example, difficult economic times often result in reduced contributions to artistic, charitable, and religious groups. Poor business earnings and declining real wages reduce tax revenues to governments, and high unemployment demands greater social expenditures by these same governments. Organizations are continually being asked to do more with less, to work in different ways, to be more responsive and caring, to provide better service, and so on. Those organizations that cannot keep up are left behind. Other businesses and countries leapfrog those that were previously leading.

Organizations such as AT&T, General Motors, and General Electric were dominant players in the late twentieth century. However, the economic impact of these historical business leaders is now dwarfed by the technology leaders of the twenty-first century: Google, Amazon, Facebook, and Apple, to name a few. As most advertisements for financial advice read: "Historical success does not guarantee future results."

The strong economic position of the U.S. for decades was in large part due to its geographic location, being physically isolated from Europe and Asia. After World War II, the U.S. dominated the world's manufacturing capacity, being the only world economic power that did not suffer significant destruction during the war. The significant prewar economies of Germany and Japan were in shambles, and those of the United Kingdom, France, Italy, and many others suffered a great deal of damage. Over the years since 1945, these countries have regained their competitive edge, and developing countries are becoming players in the world market. Japan became a major player in the global economy in the 1980s, especially in automobiles and electronics.

More recently, India has become a global leader in information technology, and China in manufacturing. The term "BRIC economies" is commonly used to refer to the growing economies of Brazil, Russia, India, and China. The obvious result of these changes is that a healthy economy, abundant jobs, high wages, and the comfortable lifestyle desired by most people of the world cannot be taken for granted; they must be fought for and earned! So what should we do?

WE NOW HAVE TWO JOBS: A MODEL FOR BUSINESS IMPROVEMENT

We used to have only one job—to do our work. We came to work, did our job, provided a product or a service, and our work was complete. There was no need to change how we did things because there was little competition. No one else was doing things differently. To survive and prosper in this new economic era, we have to make some changes. Now we must accept a second job—improving how we do our work.

Having two jobs means that we each must work to improve our personal knowledge and skills and how we do our jobs as well as get our daily work done. Managers must lead, plan, and manage how the organization can improve its performance as well as operate its day-to-day processes effectively and efficiently. This was illustrated in the billing scenario in the previous section, when Kowalski and Sons needed to improve the billing process to keep its current customers.

Organized team sports provide an excellent analogy to business because team sports operate in a competitive environment, have well-defined rules, use teamwork to succeed, and have clear measures of success (winning scores) that are monitored regularly. (We will present a statistical thinking case study involving a soccer team in Chapter 3.) The dual focus on "doing" and "improving" activities can be seen clearly in sports. For example, the "doing" work of baseball is playing the game itself. Professional baseball teams play 162 regular season games per year. But the work activities of baseball go way beyond showing up for each game, playing nine innings, and going home. The "improving" work of baseball is building individual and team skills.

The improvement cycle begins with spring training, where players get in shape and hone their skills. Players work on improving their hitting, running, and pitching. Pitchers work on controlling the curve ball, learning to throw a knuckle ball, and developing pitches they did not have before. Hitters work on hitting the curve ball or fast ball and other aspects of hitting. This work on improvement goes on all year: before the game, after the game, in the bullpen, viewing videotapes of pitching and hitting, and so on.

In the off-season, improvement activities involve weight training to build strength and speed or playing winter baseball. Coaches frequently state that star performers are not necessarily the most naturally talented but typically are those who work the hardest at improving their game. Interestingly, statistical analysis has become a key aspect of improvement, not only in baseball, but in virtually all sports. Sports analytics was popularized in the 2011 movie *Moneyball*, starring Brad Pitt, and has been growing ever since.

Figure 1.3 shows that the amount of time and effort we spend on improving how we work will increase in the future. We will also be doing more work in the future, as depicted by the larger pie on the right side of Figure 1.3. Increasing the rate of improvement is key. If the competition is also improving, the organizations that succeed will be those with the fastest rate of improvement. It is likely that Kowalski and Sons' competitors are also improving; hence, they cannot view the improvements to the billing process as a one-time event but must make improvement part of the job. Companies must continually improve or go out of business.

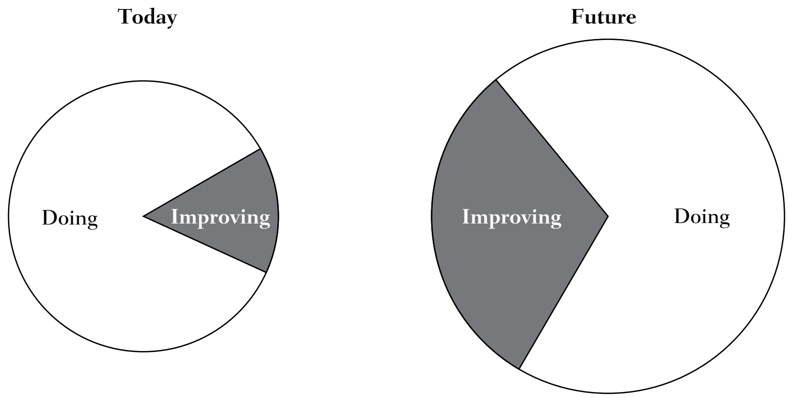


FIGURE 1.3 We Have Two Jobs: Doing and Improving

Government, health care, and nonprofit organizations also operate in this competitive environment. For example, states compete with one another and with foreign countries for investment from business and industry, which creates new jobs. States that can offer businesses the best-educated workforce and the best infrastructure (transportation, communication, etc.) at the lowest cost (taxes and regulations) tend to get new investments and jobs. The goal for all types of organizations must therefore be to improve faster than their competition.

Figure 1.4 depicts an overall model for business improvement. The doing activity is represented by the Business Process shown at the top. A series of activities, each with its own inputs and outputs, are done as a sequence of steps to produce the desired output for the customer. For example, Kowalski and Sons went through several processing steps to send out their monthly bills. The purpose, or aim, of the process is to provide a product or service of value to the customer. Note that a customer need not be someone outside the organization who purchases a product or service. A *customer* is anyone who uses the output

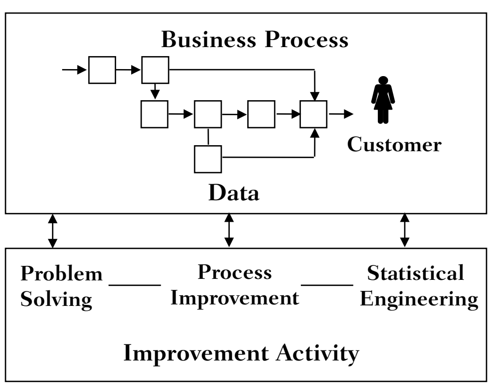


FIGURE 1.4 Improvement Model

of the process, whether within or outside the organization. Internal customers, the employees, are the key customers of the payroll process.

The improving activity is shown at the bottom of the figure. There are many different approaches to improvement, but we will focus on three types of improvement: problem-solving, process improvement, and statistical engineering.

Problem solving addresses specific problems that are not part of the normal behavior of the process. In other words, something went wrong and needs to be fixed. These issues are often resolved without fundamentally changing the process. For example, if only one customer's bill was a problem at Kowalski and Sons, they would investigate what happened to that particular bill rather than change the whole billing process. Problem solving usually involves significantly less time and cost investment than that required for true process improvement. The basic problem-solving steps are:

Step 1. Document the scope of the problem.

Step 2. Identify the root causes.

Step 3. Select, implement, and standardize corrections.

Process improvement is a series of activities aimed at fundamentally improving the performance of the process, that is, taking it to a new and higher level of performance. Some typical process improvement activities are:

- Flowcharting the process to understand it better
- Collecting data to assess the current performance
- Identifying areas where the process could be fundamentally improved
- Changing the process to implement improvement ideas
- Checking the impact of improvement efforts
- Making the improvements part of the standard way of doing business

Some problems faced in business are large, complex, and unstructured. In other words, they may just look like a "mess." There may not be a clear understanding of what the problem actually is, or how to approach it. In such cases, we need a third approach, called *statistical engineering*. This approach takes longer, because it requires significant effort just to define and structure the problem, and then "engineer" a strategy to go about solving it. A company may go through the problem-solving, process improvement, or statistical engineering cycles several times in the course of improving a process. The strategies and tools needed in these improvement approaches will be discussed in greater detail in Chapters 5 and 6. Kowalski and Sons used the process improvement model, which can require significant time and effort. If the process needs to be completely redesigned from scratch, the redesign activity is often called *reengineering*.[1]

Data are the connectors or links between the doing and improving activities. Data fuel all three of the improvement activities and increase their effectiveness. Data help us document process performance, identify problems, and evaluate the impact of proposed solutions. This was certainly the case at Kowalski and Sons. But *data* is not synonymous with *information*. For example, we presented average times for bills to be sent out, but the actual time varies from bill to bill. How should we interpret this variation? Customers do not care about average time; they only care about their bill. Therefore, we need both theoretical understanding and practical experience to properly translate these data into actionable information. A thorough conceptual understanding of statistical thinking provides us with the theoretical understanding we need, and a personal project (see Chapter 3) will help provide the experience.

NEW IMPROVEMENT APPROACHES REQUIRE STATISTICAL THINKING

New demands to improve have created the need for new management approaches, and a wide range of approaches on how to change have been proposed. These approaches typically incorporate one or more of the types of improvement discussed above; problem solving, process improvement, and statistical engineering. Among the more popular approaches today are:

- Six Sigma
- Lean Manufacturing
- Holistic Improvement
- Big Data Analytics, Data Science, and Artificial Intelligence (AI)
- The Internet of Things
- Industry 4.0

Six Sigma is a business improvement approach that seeks to find and eliminate causes of mistakes or defects in business processes.[2] Six Sigma is a statistical term that roughly translates to only 3.4 defects per million opportunities. The Six Sigma approach emphasizes understanding and documenting the business process, developing metrics and hard data, and reducing variation. This approach uses a breakthrough strategy that consists of four process improvement phases: Measure, analyze, improve, and control. The goal is to improve the process in such a way that customer satisfaction increases and there is a positive impact on the bottom line.

The Six Sigma approach was originally pioneered in 1987 by Motorola, which focused primarily on manufacturing, and was later applied by other companies including Allied Signal and General Electric (GE), which broadened the approach to include general business activities, such as financial services. Use of the Six Sigma approach expanded rapidly in the United States and around the world.

The Six Sigma methodology continues to be developed. In the late 1990s, GE added the project definition phase to the methodology, creating DMAIC (Define, Measure, Analyze, Improve, and Control). In the early 2000s, Lean Six Sigma was created by adding Lean Manufacturing (see below) concepts, methods, and tools to more effectively improve the flow of information and materials through the process, thereby increasing process speed.[3]

Lean Enterprise, originally known as Lean Manufacturing, has its roots in the Toyota Production System.[4] Lean Enterprise can be defined in many ways, although Womack and Jones, who wrote one of the classic texts on Lean,[5] emphasize Lean as an approach to eliminate waste of any type, such as excess floor space, excess inventory, excess work in progress, wasted capital, wasted effort, and wasted time. The overall approach to eliminating waste seeks to clearly identify the value that is being created, line up all activities that produce this value along a value stream, and then make the flow of the value stream a pull system instead of a push system. A pull system is one in which production is pulled, or initiated by customer orders; conversely, in a push system, products are produced as rapidly as possible and the organization attempts to push the product now in inventory to potential customers.

Womack and Jones identify five key principles of Lean:[6]

- Specify *value* by specific product
- Identify the *value stream* for each product
- Make value *flow* without interruptions
- Let the customer *pull* value from the producer
- Pursue perfection

Although its roots are in manufacturing, Lean principles have been applied to such diverse areas as banking, healthcare, and education, producing Lean Enterprise. George[7] was an early proponent of the integration of Lean and Six Sigma, to create Lean Six Sigma.

Holistic Improvement. While modern improvement approaches such as Lean and Six Sigma have proven their effectiveness on a global basis, they do have limitations. Each improvement methodology was designed with a specific type of problem in mind. Unfortunately, we have found a tendency for organizations to force fit them to problems for which they are not well suited. This is similar to the old saying, "If all you have is a hammer, every problem looks like a nail." Similarly, in improvement we should obviously fit the improvement approach to the problem, rather than force the problem to fit our predetermined improvement approach.

A more holistic approach to improvement is needed, with a broader view of how to improve business performance and a deeper understanding of the various approaches to improvement. Ultimately, a holistic approach helps develop a culture of improvement throughout the organization, including using improvement approaches as a tool for leadership development.

We define holistic improvement as "An improvement system that can successfully create and sustain significant improvements of any type, in any culture, for any business."[8] The suggested holistic improvement approach incorporates two major elements:

1. Integration of a wide set of improvement methodologies so that the most appropriate method can be used to attack a given problem. In other words, no single methodology can solve the full breadth of the problems noted; therefore, multiple methodologies must be integrated. As noted above, the problem should determine the methodology, not vice versa.
2. A deployment system that provides the needed infrastructure required to create and sustain improvement across the spectrum of the problems. By infrastructure, we mean the business processes and organization needed to sustain improvement. This includes defined roles, planning systems, budgeting processes, formal project reviews, and so on, just as organizations typically provide to manage the finance organization.

We return to holistic improvement in Chapter 5, where we provide three distinct improvement frameworks to utilize for problem solving, process improvement, and statistical engineering.

Big Data Analytics, Data Science, and Artificial Intelligence (AI). Data are being collected at an ever-increasing pace, through social media, online transactions, and scientific research. According to IBM, 1.6 zettabytes (10^{21} bytes) of digital data are now available. That's a lot of data, enough to watch high-definition TV for 47,000 years.[9] Hardware, software, and statistical technologies to process, store, and analyze this data deluge have also advanced, creating new opportunities for analytics. Various terms have been used to describe these technologies, such as Big Data analytics, machine learning, data science, and AI. Modern technology companies, such as Google, Amazon, and Facebook have been largely built upon these methodologies.

While these terms are sometimes used interchangeably, they do have unique definitions, and should not be used as generic "buzzwords." Big Data analytics simply means analysis of very large data sets, those too large to analyze with traditional hardware and software. Google and Amazon excel at this. Data science is also a general term referring to the acquisition, processing, and analysis of data, typically with a strong focus on the underlying computer science and coding. Machine learning, upon which we elaborate below, consists of a set of modern modeling methods that are computationally intensive and focus on accurate prediction, without necessarily developing an understanding of the cause and effect relationships among the variables. AI refers to actions or decisions made by a computer that, if a human made them, we would consider the human to be intelligent. AI typically involves automated equipment, such as self-driving cars, or automated decisions such as approving credit card applications.

To elaborate on machine learning, many new modeling methods (spanning statistics, applied mathematics, and computer science) have been developed in the past 30 or so years to take advantage of the large amounts of data now available. Methods such as neural networks, classification and regression trees (CART), and support vector machines (SVM) are now commonly used, and are generally referred to as machine learning. "Deep learning," now a popular term in the scientific and popular media, refers to multilayer (more complex) neural networks. Recently, ensemble methods that resample the data and integrate multiple models into an overall grand model, such as "random forests," have become more popular.[10] Ensemble methods have won many of the modeling competitions so common today, such as the Netflix $1,000,000 prize, as well as the competitions available on Kaggle.com.

We discussed the *Internet of Things* previously. As noted, physical devices are now becoming connected through the Internet, often bypassing humans. This allows engines, motors, robots, and so on, to share data and make decisions. When they do so without human involvement, this would be considered AI. The Internet of Things and Big Data create the foundation for *Industry 4.0*, involving cyber-physical systems. As discussed previously, such systems, Uber being one example, digitally integrate humans, data, and physical equipment.

Each of these approaches and philosophies is useful, and the best aspects of each can be integrated with the management approach an organization is currently using. The result is a new management approach that helps the organization better serve the needs of its customers and compete effectively in the marketplace. Three common themes run through these management approaches:

1. Viewing work as a process
2. Using data to guide decisions
3. Responding wisely to variation

These three items are part of the body of knowledge known as statistical thinking. This body of knowledge and its associated skills are essential to the successful management and improvement of any business. Statistical thinking is a philosophy of learning and action based on these fundamental principles:[11]

- All work occurs in a system of interconnected processes.
- Variation exists in all processes.
- Understanding and reducing variation are keys to success.

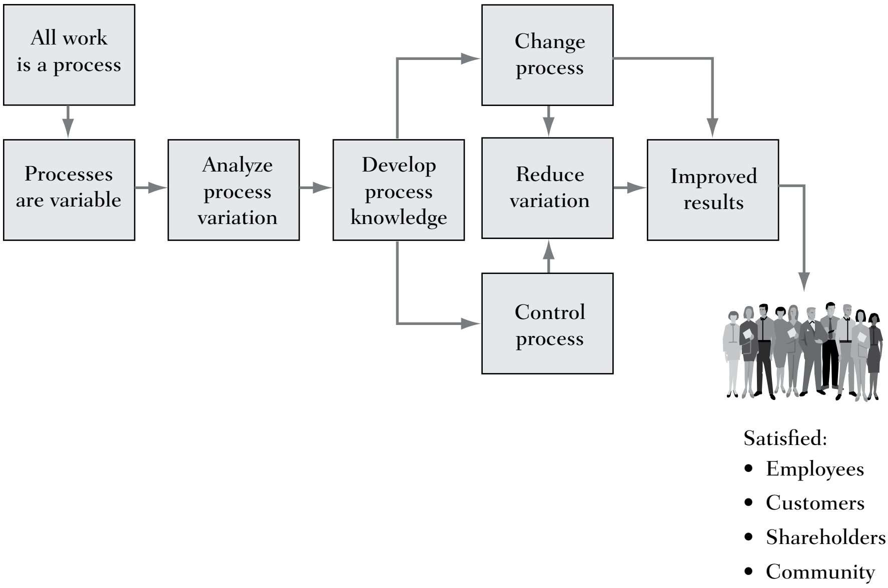


FIGURE 1.5 Steps in Implementing Statistical Thinking

These principles work together to create the power of statistical thinking. The steps in implementing statistical thinking are shown in Figure 1.5. We begin by recognizing that all work is a process and all processes are variable. We must analyze the process variation to develop knowledge of the process. You cannot improve a process that you do not understand. Note that these core principles are similar to the common themes of recent management improvement efforts presented earlier. With knowledge of the process, we are in a position to act to improve that process.

From a statistical point of view, improvement activity—including statistical engineering, process improvement, and problem solving—can be viewed as working on either of two process characteristics: (1) reducing variation through tighter control of the process or (2) improving the overall level (average value) by changing the process target, which may also result in reduced variation. For example, the primary objective of Kowalski and Sons' billing efforts was to reduce the average time to get bills out. They also wanted to reduce the variation from bill to bill. The end result of using statistical thinking is business performance that satisfies the stakeholders: customers, employees, the community in which the business operates, and the shareholders.

The terms *average* and *variation* are critical to applying statistical thinking. For most processes the average is the central value around which the process varies. Variation results when two or more measures of the process are different, which is the rule rather than the exception. The calculation and use of average and variation statistics are discussed in Chapter 2.

PRINCIPLES OF STATISTICAL THINKING

The first principle of statistical thinking is that *all work occurs in a system of interconnected processes*. This principle provides the context for understanding the organization, improvement potential, and sources of variation mentioned in the second and third principles. A process is one or more connected activities in which *inputs* are transformed into *outputs* for a specific purpose.

This is illustrated in Figure 1.6. For example, mailing or electronically submitting bills requires that records are kept on charges (inputs). These records must be processed (aggregated for a month, reduced by payments made, checked for accuracy and applicable discounts, and so on), often with the aid of automated computer systems, into a monthly bill. Any discrepancies or errors must be resolved, often through a manual process. These bills must then be physically printed and stuffed into the appropriately addressed envelopes and delivered to the desired mailing system (U.S. Postal Service, Federal Express, United Parcel Service, etc.) or posted to appropriate electronic addresses. The bill a customer receives is the output of the process.

For a simpler example, the series of activities one goes through to get to work or school in the morning (getting out of bed, bathing, dressing, eating, and driving or riding) can be thought of as a process. Any activity in which a change of state takes place is a process, as depicted in Figure 1.6.

We encourage people to "blame the process, not the people" when working on improvement. Joseph M. Juran pointed out that the source of most problems is in the process we use to do our work. He discovered the "85/15 rule," which states that 85% of the problems are in the process and the remaining 15% are due to the people who operate the process. W. Edwards Deming states that the true figure is more like 96/4.[12] We may debate the correct figure, but it is clear that the vast majority of problems are in the process. Businesses

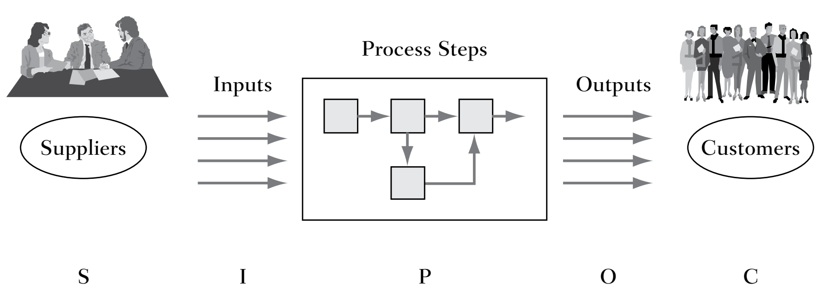


FIGURE 1.6 SIPOC: A Process View of Work

and other organizations are made up of a collection of processes. Business processes interconnect and interact to form a system that typically provides a product or service for a customer (the person who receives the output of the system). Viewing business as a collection of processes is discussed in detail in Chapter 4. Some typical business processes are:

- Customer service
- Business planning
- New product development
- Employee recruiting and orientation
- Maintaining the company IT network
- Manufacturing
- Equipment procurement
- Patent application
- Accounting
- Laboratory measurement

The second principle of statistical thinking is that *variation exists in all processes*. This provides the focus for improvement work. Variation is the key. If there were no variation:

- Processes would run better.
- Products would have the desired quality.
- Service would be more consistent.
- Managers would manage better.

Focusing on variation is a key strategy to improve performance. Of course, in certain situations, variation is desirable, such as wanting variety in menu items in a restaurant, valuing diversity among team members, and so on. Intended or desirable variation is valuable and should be promoted. It is unintended variation that we will focus on here.

Variation is a fact of life. Variation is all around us. It is present in everything we do, in all the processes we operate, and in all the systems we create. Variation results when two or more things, which we may think are exactly the same, turn out to be different. Some examples are:

- Restaurant service time varies from day to day and from customer to customer.
- Tires wear at different rates.
- Tomatoes of the same variety vary in weight.
- Shirts of the same size fit differently.
- Cars of the same model perform differently.

To understand and improve a process we must take variation into account. Variation creates the need for statistical thinking. If there were no variation, there would be little need to study and use statistical thinking.

The third principle of statistical thinking is that *understanding and reducing variation are keys to success*. The focus is on unintended variation and how it is analyzed to improve performance. First, we must identify, characterize, and quantify variation to understand both the variation and the process that produced it. With this knowledge we work to change the process (e.g., operate it more consistently) to reduce its variation.

The average performance of any process (e.g., average time in days to get bills out, average waiting time in minutes to be served in a restaurant, or average pounds of waste of a printing process) is a function of various factors involved in the operation of the process. When we understand the variation in the output of the process, we can determine which factors within the process influence the average performance. We can then attempt to modify these factors to move the average to a more desirable level.

Businesses, as well as customers, are interested in the variation of the process output around its average value. Typically, consistency of product and service is a key customer requirement. For example, customers of a shipping company do not want shipments to arrive sometimes in a day and other times in a week. Such variation would make it difficult to plan when to actually ship goods. Similarly, in accounts payable, we do not want to pay bills too late because we may lose discounted terms (often a 2% discount for prompt payment). If we pay too early, however, we lose interest on the cash. If we can reduce the variation in the process, we can consistently pay right at the due date and receive discounted terms and minimize loss of interest.

This results in many improvement efforts that focus on reducing variation, for example, reducing restaurant waiting time from 0 to 10 minutes to 0 to 5 minutes, reducing the variation in real estate sales from 20 to 40 units/month to 25 to 30 units/month, or reducing bill mailing time from 10 ± 4 days to 10 ± 1 days. Low variation is important to customers, and they will sometimes accept less desirable "average performance" to obtain better consistency.

Three types of variation that we may need to reduce are *systematic variation*, which is caused by specific inputs and process variables, *outliers* due to "special causes" or unforeseen events, and *random* or "common cause" variation that is neither explainable nor predictable, except within limits. Systematic variation is variation in outputs caused by variation in inputs or process variables. It can generally be explained, and predicted. For example, suppose Karen is an Uber driver in a major city. She typically begins her shift at 5:00 p.m. because she has a second job during the day. Today, however, she was off from her second job, and so starts her Uber shift at 11:00 a.m. She knows that she will not have as large a demand for rides at 11:00 a.m. as

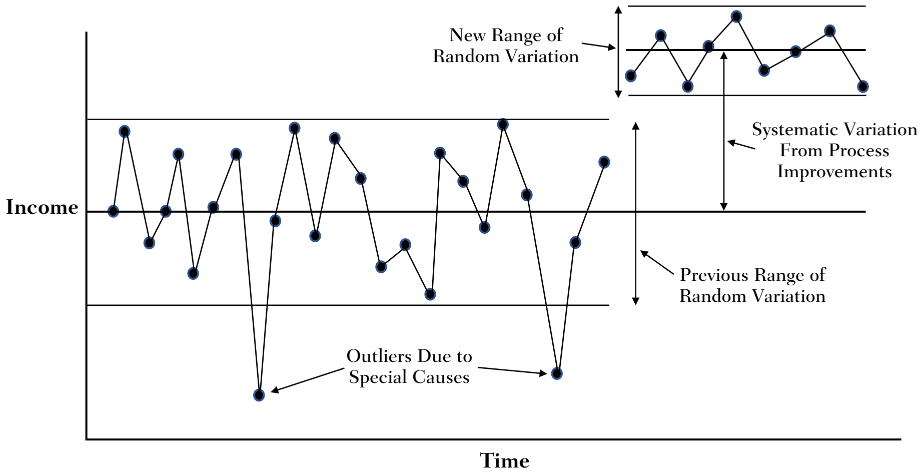


FIGURE 1.7 Systematic, Random, and Outlier Variation

at 5:00 p.m. because it is not during rush hour. She knows why her income per hour will be lower that day—because of variation in demand at different times of the day. Further, Karen can predict that this will be the case next week as well. Hourly work patterns are a process variable impacting the income of Uber drivers (the output) that can be explained and predicted. It is systematic variation. Much of statistical analysis is aimed at understanding and modeling the impact of systematic variation.

Outliers are extreme data points that don't fit the pattern of the rest of the data, as shown in Figure 1.7. Outliers are often invalid data points due to a measurement error, but may also represent legitimate data points. They should not be automatically eliminated from data sets – they might be telling you something! Outliers that are legitimate data points are typically due to *special causes* that are outside the normal or typical behavior of a process. They can usually be explained but can't be predicted in advance (if they could be predicted, we would consider it systematic variation). Going back to Karen the Uber driver; suppose Karen's car breaks down at 2:00 p.m. and has to be towed to a garage. Karen is not likely to make much money sitting in a garage, waiting for her car to be repaired.

Her income per hour for the day will likely be an outlier, much lower than for a typical day. However, this is a "valid" data point, in that this is actually how much she made that day; it is not a mistake. Karen understands why her income was down, i.e., this outlier variation is explainable, but of course she couldn't predict that her car would break down today. In many cases we can identify and eliminate the special cause, reducing the likelihood of similar issues in the future. Eliminating special causes of outliers stabilizes the process by reducing outlier variation. For example, going forward, she can make sure that she does all the preventive maintenance on her car recommended by the manufacturer. Of course, eliminating causes of outliers only returns Karen to her normal income (unless she experiences a "positive" outlier); it does not fundamentally improve it.

Normal or typical variation is called *random variation*, or *common-cause variation*. Random variation appears to be just that; random. We cannot generally explain or predict random variation. For example, we can't explain why we obtained a 7 on a roll of two dice, and then a 10 the next time. Rolling dice is a random process. Similarly, Karen may find that she makes between $125 and $200 on a typical shift. One day she makes $150, and then the next day, also a weekday at the same hours, she makes $185. Why? These values are both typical, so not qualifying as outliers. However, she can't identify any inputs or process variables that explain or predict the $185 versus $150. Therefore, this is random variation. The distinction between these three types of variation will be discussed in more detail in Chapter 3.

A key contribution of statistical analysis of data is to segregate variation into these categories. Understanding the distinction between the three major types of variation is critical, because the best approach to addressing them is different for each. Since outliers are atypical, we can often eliminate the root special causes without fundamentally changing the process. Maintaining her car wouldn't cause Karen to either quit her job or modify her working schedule. In general, using a problem-solving approach we identify what was different in the process when it produced the unusual result. There should be a specific, identifiable reason for the outliers. Even if we eliminate the causes for these points through problem solving, however, we are left with the normal level of random variation.

Making further improvement typically requires studying the process as a whole, because there are no unusual results (outliers) to investigate. In other words, there is no single, identifiable reason income was high one day and low the next. Using a process improvement approach, we study normal process variation and try to discover the input and process factors that are the largest sources of variation. That is, we try to identify some specific causes of what we originally thought was random variation. These causes would be considered systematic sources of variation. A portion, perhaps large, perhaps small, of the remaining "random" variation may actually be predictable. Once these systematic sources of variation are identified, typically inputs or process variables, we may be able to utilize them to fundamentally change the system; that is, we may be able to improve the overall process.

In Karen's case, if she collects and analyzes data carefully, she may discover that her income does in fact go up or down due to predictable variables. These could include the weather (time of year), the day of the week, the presence of major sports or entertainment events in town, and so on. Using this information, she might modify her shift schedule, that is, change her process to fundamentally improve it. She will always have some random variation, but over time she may be able to identify and utilize more systematic sources of variation to improve her income. The key point is that breaking variation down into these three categories is a critical aspect of statistical thinking, and of improvement in general. All variation is NOT created equal.

APPLICATIONS OF STATISTICAL THINKING

Statistical thinking can be used in all parts of an organization and in all job functions. For example, the job of the executive is to lead the organization in a common direction. The executive must gain agreement on the desired direction and get employees to align all their activities with this common direction. The result of this reduced variation in activity is less wasted effort, less rework, reduced costs, and faster improvement of organizational performance. An engineer in a factory may want to reduce the variation in a product characteristic to meet customers' specifications a higher percentage of the time. This may require using a better process control procedure or developing a deeper understanding of the process so that fundamental changes can be made to better operate the process.

A business analyst may wish to find ways to change the "closing" process (calculating final profit and loss for a specific time period, such as quarterly or annually) so that, on average, the accounting books are "closed" in 2 days rather than the current 15-day average, thereby reducing overtime work required and increasing customer satisfaction. Customers in this case would include corporate management and the financial community. This objective will likely require a better understanding of how the closing process works and what persons and groups are involved. Armed with this knowledge, we can find ways to improve the process to meet business needs.

A restaurant manager may want to understand why it is taking a long time to serve some customers. Data on customer service time is needed to document current performance and to determine whether newly introduced procedures are working. Properly collected data are also useful in identifying causes of poor performance and possible solutions.

All of the problems identified in these examples can be addressed using statistical thinking. As noted earlier, statistical thinking also applies to various kinds of work in all organizations and functions, including:

Type of Organization	**Type of Function**
Manufacturing	Marketing
Financial services	Sales
Education	Manufacturing
Government	Research and development
Health care	Engineering
Retail sales	Human resources
Transportation	Information systems
Software	Purchasing
Restaurants	Finance

Statistical thinking is also useful in the leadership and management of organizations. We noted earlier that the job of the executive is to move the organization in a common direction. To do this, the executive must gain agreement among the employees regarding the direction the organization should go. This will reduce variation in employee actions.

Similarly, many organizations have reduced the number of their suppliers. This action is based on the fact that a large number of suppliers increases the variation in incoming supplies, which in turn increases the variation in the output of the organization. Customers want consistent products and services. When a company decreases the number of suppliers, the company generally decreases the variation in their product or service.

Any time you improve a process, product, or management system, you are requiring the involved employees to change how they work and the design of the processes they use to do their work. Such actions require a leader to help the organization make the needed changes. Such leaders can be anywhere in the organization, including scientists, engineers, financial analysts, managers, and others. Statisticians and quality professionals often serve in leadership roles as well.[13]

SUMMARY AND LOOKING FORWARD

In this chapter, we have discussed today's business realities and the need for businesses to improve, a model for business improvement, some new management approaches that have been proposed, and principles and applications of statistical thinking. The following is a list of the most important ideas and conclusions addressed in this discussion:

- Global competition and rapid technology change have forced companies and organizations around the world to change.
- Improvement is needed for an organization to survive and prosper.

- We now have two jobs: to do our work and to improve on how we do our work.
- Globally, systems of improvement are changing, and many new approaches have been proposed.
- Statistical thinking is an integral part of the common themes that run through these new approaches.
- Statistical thinking is based on three principles: All work occurs in a system of interconnected processes, variation exists in all processes, and understanding and reducing variation are keys to success.
- A key aspect of statistical thinking is breaking overall variation down into systematic, random, and outlier variation.
- Broad use of statistical thinking can help an organization improve operations and its management system.

EXERCISES: CHAPTER 1

1. Update the data on the U.S. balance of trade and the U.S. median hourly earnings (Figures 1.1 and 1.2). Construct a plot of these measures of the U.S. economy versus time, and comment on the trends you see, based on the latest information. Comment on how global events may have impacted these trends.
2. Describe how global competition and its effects on the economy have affected you and your family *personally*. Consider the effects of global competition on all aspects of society (education, health care, government, religious institutions, and so on). Be sure to consider how these effects link together. For example, how do effects on business impact non-profits and government?
3. Describe a work task, a hobby, or another activity you regularly do, and list sequentially the various things you do to complete this activity. Discuss the complexity of your list; why are this many steps required to complete the activity?
4. Describe a specific activity (work or play) of particular interest to you. Explain how you would go about improving this activity. Differentiate between doing and improving activities.
5. Understanding variation is at the core of statistical thinking. Describe how variation has affected you in your daily life. Give at least three specific examples.
6. Using Figure 1.5 as a model, explain how statistical thinking can be used to improve the situation you described in Exercise 3.
7. Discuss the impact that social media is having on the world business climate. How might social media provide new opportunities for applications of statistical thinking?
8. Using one of the improvement approaches discussed in the section "New Improvement Approaches Require Statistical Thinking," describe how that approach can improve organizational performance and how statistical thinking is an integral part of the approach. Use the Internet or other sources to find background information on, and examples of, the chosen management method.
9. Do you feel that statistical thinking can aid in the implementation of artificial intelligence (AI)? Why or why not?
10. Organizations are accumulating vast amounts of data today through the Internet (Big Data). Identify specific ways the Internet can be used to obtain information and data to improve business processes. (Hint: Note the elements of business processes and the various types of improvements—such as reducing errors and decreasing cycle time—and approaches to improvement.)

NOTES

1. M. Hammer and J. Champy, *Reengineering the Corporation* (New York: HarperCollins, 1993).
2. R. D. Snee, and R. W. Hoerl, *Leading Holistic Improvement with Lean Six Sigma 2.0*, 2nd ed. (London: Pearson Education, 2018).
3. M. L. George, *Lean Six Sigma: Combining Six Sigma Quality with Lean Speed* (New York: McGraw-Hill, 2002).
4. Y. Monden, *Toyota Production System: An Integrated Approach to Just-in-Time*, 4th ed. (Boca Raton, FL: CRC Press, 2012).
5. George, *Lean Six Sigma*.
6. Ibid.
7. Ibid.
8. Snee and Hoerl, *Leading Holistic Improvement with Lean Six Sigma 2.0*.
9. M. Ebbers, "5 Things to Know About Big Data in Motion," *IBM developerWorks Blog*, (2013). Available at: https://www.ibm.com/developerworks/community/blogs/5things/entry/5_things_to_know_about_big_data_in_motion?lang=en.
10. G. D. James, T. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning* (New York: Springer, 2013).
11. *Statistics Division of the American Society for Quality Control, Glossary and Tables for Statistical Quality Control*, 3rd ed. (Milwaukee, WI: Quality Press, 1996).
12. B. L. Joiner, *Fourth General Management* (New York: McGraw-Hill, 1994), 33–34.
13. R. D. Snee, "Digging the Holistic Approach: Rethinking Business Improvement to Improve the Bottom Line," *Quality Progress,* (October 2009), 52–54.