Data Analysis Through Modeling: Thinking and Writing in Context

Kris Green and Allen Emerson

Fall 2011 Edition[1]

About this text

Data Analysis Through Modeling is a one-semester data analysis and calculus text that can be used as part of a one-, two- or three semester sequence of mathematics courses usually required of business and management undergraduate majors. We believe the following features distinguish this text from other texts in the curriculum:

⇒ Data-driven, open-ended problems
⇒ Extensive use of spreadsheets throughout the text as more than just a calculator
⇒ Key problems framed as realistic business memos
⇒ Follows recommendations of MAA's Curriculum Foundations Project CRAFTY report for business and management

The increasingly information-driven demands of business in the 21st century require a different emphasis in the quantitative skills and ways of thinking than traditional mathematics courses have provided in the education of managers. This emphasis has to do with becoming comfortable in the world of data and mathematical models, being able to use technology as a tool through which to think, and expressing one's thinking effectively in writing.

The key, we believe, is data analysis through modeling. Data analysis for us means "What can we find out about this data set relevant to our problem?" Models for us are such things as: averages, boxplots, histograms, single- and multivariable regression equations, both linear and nonlinear. These models are proxies for data that are too complex to understand any other way. We think of calculus as a way of analyzing certain kinds of models, which in turn, reveals something about underlying data structures. Our treatment of calculus emphasizes basic concepts, such as rates of change, constrained optimization, and interpretations of area under a graph, and their applications to business problems. We use spreadsheets to develop numerical methods for both differentiation and integration while deemphasizing symbolic manipulation. We use Excel's Solver routine instead of the simplex method to solve linear programming problems. Using Solver has the advantage that we can also solve nonlinear programs.

As we developed this text, we found the introduction of spreadsheet technology for analysis of data not only changed our teaching approach and the content of the course, but it caused us to modify our assignments as well. We found that we simply could not get the quality and depth of understanding we desired from our students by using conventional exercises. We found that students have to explain their thinking and make explicit their assumptions and inferences. In short, we had to supplement our more conventional exercises with memoranda problems with accompanying data files that students respond to in an appropriate business format that are, in turn, read by their supervisor. Further, we find that students learn more by having a chance to revise their work based on instructor/supervisor feedback. All of which should give an indication as to why the book is subtitled "Thinking and Writing in Context."

Although the text has a unit of descriptive statistics and develops regression all the way through multivariable regression with interaction terms, Data Analysis Through Modeling is not a statistics text. Most one-semester introductory statistics courses do not treat regression at the level presented in this text. Moreover, most introductory statistics texts do not give

the same emphasis to descriptive statistics that this text does, which is to use these relatively simple concepts for rather deep analysis. *Data Analysis Through Modeling* fits well with an introductory statistics course that primarily deals with probability and hypothesis testing.

How this text fits into the curriculum

We recommend the following tracks for a three-credit-hour, semester-long course using *Data Analysis Through Modeling*:

- For students not having a prior statistics course: Chapters 1-9, 11-12 [11 chapters]. This course would not contain calculus and would be the first in either a two- or three semester sequence: 1) data analysis and statistics or 2) data analysis, statistics, and calculus. In our experience, students then do quite well in the follow-up statistics course after their experience with our approach to data analysis.

- With a statistics prerequisite: Chapters 1-3, 7-9, 11-17 [12 chapters]. This course would contain calculus and constitute the second course in a two-semester sequence containing probability and hypothesis testing, data analysis, and calculus.

The basic concepts of calculus are emphasized and applied to business problems involving marginal analysis, optimization and area under a curve. As recommended by CRAFTY, formal techniques of symbolic manipulation are kept to a minimum, whereas spreadsheets are used extensively not only for finding numerical solutions but, equally important, for the development of the basic concepts of calculus themselves.

The Technology Used in this Text

In addition to problem solving in the dynamic environment of spreadsheets, students will have the opportunity to learn about and use the following Excel tools: pivot tables, sorting, stacking and unstacking data, basic statistical functions, frequency tables, sumproduct, building boxplots and histograms, correlation tables, simple regression, multivariable regression (quantitative and qualitative), scatterplots, trendlines, Goal Seek, SOLVER table and graphing in three dimensions. In addition, students will develop many basic computer literacy abilities, such as copying and pasting and integrating numerical, textual and graphical analyses into a single Word document. But what is most important about the way students learn these tools is that they are all taught in the context of solving business-type problems; this context, we believe, is critical for students learning how to transform these tools from a set of instructions to follow into a method of thinking and analyzing data.

The Structure of the Book

This text is organized into five units, not all of which can be covered in one semester, as mentioned above. The chapters in each unit are all connected through a common "thinking

Unit	Thinking Strategy
Quantifying the World	Students learn the importance of data and how to locate data in real world situations.
Analyzing Data Through Spatial Models	Students learn how to use basic charts and graphs to deeply understand a problem situation.
Analyzing Data Through Linear Models	Students learn how to apply proportional reasoning to understand data with one or more independent variables.
Analyzing Data Through Nonlinear Models	Students learn to build models by linearizing non-proportional data and learn how to interpret these in realistic situations.
Analyzing Data Through Calculus Models	Now that students understand how to build models from data, they learn how to use concepts from calculus to understand the problem from which the data and the model were derived.

Table 1: Units and thinking strategies covered in the text.

strategy". The thinking strategies are described in the table 1. The breakdown of topics in each chapter within the units is described later.

Each chapter is designed to be covered in one week of a typical semester course. Since the homework problems (see below) come at the end of a chapter, the homework schedule should, ideally, consist of one assignment per week. Each chapter's introduction provides a brief overview. It also includes a list of goals and objectives that the student should have after completing the chapter. After the introduction and overview, the main content of each chapter is separated into two major sections, each of which consists of the following:

Discussion. This presents a short overview of the chapter or discusses a short motivational example illustrating the use of the chapter material. The material in this section is conceptual in nature.

Definitions and Formulas. This lists the factual information of the chapter in the form of definitions, formulas, graphs, and methods of computing. It is intended as a reference guide.

Worked Examples. These offer worked examples of using the formulas and techniques of the chapter. This material is more often procedural in nature, but uses concepts to unpack and apply the material to realistic situations from the business world.

Explorations. These involve small scenarios, often supplemented with data in Excel. They are open-ended and require discussion and scaffolding. These are basically guided-discovery type activities and are ready-made in-class activities, but can also cbe completed by students outside of class in order to enhance their understanding of the chapter material.

How to Guides. These offer the details for getting Word, Excel and StatPro (an Excel add-in) to handle the computations and graphing needed to complete the exercises.

Homework Problems

Each chapter within a unit is designed to provide the material for a weekly homework assignment at the end of the second section of the chapter. The problems at a chapter's end come in three types: Mechanics and Technique Problems, Application and Reasoning Problems, and Memo Problems (which include Communication and Professionalism skills). Although we consider the memos to be the heart of any course using this book, the number of memos instructors choose to assign on a weekly basis will vary and the two other types of problems work very well to provide a balanced weekly assignment load.

Mechanics and Technique Problems. These problems involve straightforward calculations by hand or, more often, with the computer, and use the basic definitions, formulas, and computer techniques from the chapter.

Application and Reasoning Problems. These problems require students to analyze data or apply the concepts of the chapter to small decision-making scenarios. Many of these require students to explain their thinking in a few short sentences so that the inferences they have drawn from the data and other information are made explicit.

Memo Problems. Each chapter concludes with a memo problem from a supervisor at Oracular Consulting. The memos are written in the style of a management memo, often having a rather open-ended feel, and will most often direct the analysis staff (the students) to analyze some data for a client, using the tools of that chapter (and possibly previous chapters). Students are expected to reply to these memos with their own professionally written memos or reports. Most memo problems usually permit more than one "correct" response. We have developed detailed "rubrics" for assessing each memo which are invaluable should the instructor choose to have students revise and resubmit their memos. These can be found in the *Instructor's Guide*. These rubrics do not contain "answers" per se, but rather statements to be checked off by the instructor that note lapses in analysis, missing pieces, incorrect or misapplied mathematical/computer procedures, or point out structural writing difficulties. These statements are divided into three discrete areas: Mechanics and Technique, Applications and Reasoning, and Communication and Professionalism, and each of these three is divided into two levels of competence, Expected and Impressive (see the appendices for an example). In the *Instructor's Guide* we describe in detail how we arrive at grades.

Entering Student Profile

As a student entering a course using this book, or as someone using this book on their own to gain new skills, techniques, and concepts about quantitative analysis in the business world, you shold have some skills in the areas of mathematics, the use of technology, and writing.

Mathematics background: Basic algebra skills are essential, but the text does not require well-honed algebraic skills as a pre-requisite. What is most essential is the abstraction that algebra supports in moving from concrete objects to expressions and functions

with parameters and variables. Students should have had a mathematics background up to, but not necessarily including, precalculus.

Technology background: The text does not assume that the students have any knowledge of spreadsheets, though in our experience most have some familiarity with computers and spreadsheets, Excel in particular.

Writing Background: In our experience, students gain the most from this text when it is taught in a writing-intensive format, using a selection of the chapter memo problems (including revisions). Most first-year college writing course requirements will have prepared students sufficiently to write at the level the memos demand.

Exiting Student Profile

By the end of a course based on this text, we expect students to have developed capabilities in three areas. The first area (mentioned above) is "Mechanics and Techniques," which includes knowledge of basic mathematical notation and symbol manipulation as well as basic technological (especially spreadsheet) skills for structuring problems for solutions. The second area is "Application and Reasoning," which covers the ability to contextualize the mathematical ideas, to extract quantitative information from a context, and to make logical inferences from quantitative analyses. The final area is "Communication and Professionalism," which covers the ability to write coherently about a problem and its proposed solution and to communicate this analysis in a professionally appropriate manner.

Specifically, a student earning an average grade in a course based on this text would have the capabilities in each of the three areas shown in the outline below.

Mechanics and Techniques

- Has had experience formulating and interpreting algebraic, graphical and numerical mathematical models
- Has used spreadsheets to apply various mathematical, statistical, and graphical tools to business situations
- Understands enough about data analytic techniques to effectively communicate with statisticians and other types of expert analysts
- Is competent and comfortable with spreadsheets
- Has learned to use technology as a tool with which to think

Application and Reasoning

- Understands how to define a problem situation in terms of data
- Understands the basic design of data collection forms and how to employ them
- Has experience in working in open-ended, ambiguous problem situations
- Understands the interpretive power of graphical displays of data
- Understands the power and limitations of mathematical models
- Has experience in interpreting the parameters and coefficients of mathematical models

- Is capable of drawing contextual inferences from statistical and graphical analysis

Communication and Professionalism

- Knows the importance of writing in the workplace
- Can write competent memos and reports as part and parcel of one's job
- Knows how to integrate and arrange statistical and graphical elements in a word processing document to produce a convincing argument
- Has learned to consider the reader's response to a memo
- Has learned to plan ahead to meet the demands of the course
- Persists when the path is not clear
- Has learned self discipline in accomplishing long and complex tasks

Some Words About Level of Difficulty

Viewed apart from a context of a memo, the mathematics, technology, and writing demands of certain chapters may not seem very difficult when taken separately. But when students analyze a data set using Excel, interpret and draw inferences from mathematical formulations within specific problem contexts and then organize the various charts, computer output, and text into a coherent and readily understood memo, they find the work to be anything but easy. Indeed, instructors of this text invariably comment on how they themselves have been challenged by the problems. The open-ended nature of the problems (e.g. see the Chapter 1 memo) contributes to this challenge, as well as the sheer amount of time it takes to complete the whole process. This is one of the reasons that instructors may not wish to assign a memo problem every week, especially when requiring revisions, which students mightily appreciate and benefit from.

Some Words About Plagiarism and Working Together

We require all memos to be submitted electronically through a course website (Blackboard) in Word. This enables us to issue the following policy that eliminates concerns about plagiarism:

"We encourage you to work together and to seek help when you need it. Our only requirement is that you write your own memo in your own words."

Invariably, two or three students will copy each other's work sometime in the beginning of the semester. Because each writer's voice comes through so strongly even in the memo genre, duplication is easy to detect. Furthermore, technology is an aid in identifying copying. For example, Microsoft Word has a feature called *compare and merge documents* (under tools) that superimposes one document upon another showing all differences in red (every space, every comma, whole chunks of text, etc.) or, more importantly, no differences. Tips on using this tool are available in the *Instructor's Guide*. Once identified, instructors can respond with the following notification: "Computer analysis shows that significant portions of your memo and Mike's memo are identical. While we encourage you to work together, we do require that you do your own write up. Friendly warning." There are no copying problems from this point on. Maybe word gets around the class about the "computer analysis."

Copyright Notice

Chapter Details

Unit 1. Quantifying the World. Students learn the importance of data and how to locate data in real world situations.		
Chapter	Content	Memo Regarding
1. Problem Solving	Framing a problem in terms of data	Performing the up-front analysis in response to a RFP from Carnivorous Cruise Lines concerning lack of attendance at its entertainment venues (No data file)
2. Understanding the Role of Data	Collecting and organizing data to support problem solving	Creating data collection forms and displaying sample test data in spreadsheets for the Carnivorous Cruise Lines RFP (Create your own data file)
3. Using Models to Interpret Data	Building simple models to analyze data using the mean, standard deviation and pivot tables	Analyzing sample data from Carnivorous Cruise Lines to make changes in the entertainment schedule (Data file)

Unit 2. Analyzing Data Through Spatial Models. Students learn how to use basic charts and graphs to deeply understand a problem situation.		
Chapter	Content	Memo Regarding
4. Box-and-Whisker Plots	Using boxplots and associated measures to build and analyze spatial models of data	Using boxplots to explore the salary structures of four different companies for two quite different managers in need of a job (Data file)
5. Histograms	Using z-scores and histograms for understanding different distributions of data	Analyzing customer wait times at a fast food restaurant in response to customer complaints of poor service (Data file)
6. Interpreting Spatial Models	Estimating statistics from summary data and connect the different spatial models (boxplots and histograms) to build a more complete understanding of a set of data	Analyzing ten different stocks in order to build financial portfolios for two quite different clients. (Data file)

Unit 3. Analyzing Data Through Linear Models. Students learn how to apply proportional reasoning to understand data with one or more independent variables.		
Chapter	Content	Memo Regarding
7. Correlation	Picturing and quantifying the relationship between two variables using correlation and trendlines	Using and interpreting trendlines to determine how in-city and out-of-city driving conditions effect maintenance costs for a trucking fleet (Data file)
8. Simple Regression	Using simple linear regression to measure the effect of one variable upon another and to interpret how well our models fit the data	Building and interpreting simple regression models regarding the how various variables affect ridership on a commuter rail system (Data file)
9. Multiple and Categorical Regression	Extending regression modeling into many dimensions and using qualitative variables	Building successive multivariable models using quantitative and qualitative variables to analyze how gender might be implicated in the salary structure at a company (Data file)
10. Is the Model Any Good?	Exploring the reliability of linear models and introducing interaction terms into the models	Developing more realistic models of the truck fleet maintenance costs using interaction terms and stepwise regression analysis (Data file)

Unit 4. Analyzing Data with Nonlinear Models. Students learn to build models by linearizing non-proportional data and learn how to interpret these in realistic situations.		
Chapter	Content	Memo Regarding
11. Graphical Approaches to Nonlinear Data	Examining a variety of nonlinear graphical models with one independent variable (logarithmic, exponential, square, square root and reciprocal) and their transformations	Analyzing various data sets from a customer who wants better models for each set than those created by Excel's trendlines; this is accomplished by shifting and scaling the basic models and computing the goodness of fit for each (Data file)
12. Modeling with Nonlinear Data	Building and interpreting nonlinear regression models, including general power models and multiplicative models in several variables	Creating and comparing multivariable models (one linear and one multiplicative) to help analyze operating costs at an insurance company (Data file)
13. Nonlinear Multivariable Models	Extending the variety of nonlinear multivariable models to include quadratic models developed from interaction terms	Developing more accurate models of the commuter rail system data by using quadratic interaction terms (Data file)

Unit 5. Analyzing Data Using Calculus Models. Now that students understand how to build models from data, they learn how to use concepts from calculus to understand the problem from which the data and the model were derived.		
Chapter	Content	Memo Regarding
14. Optimization and Analysis of Models	Using calculus (derivatives) to interpret and optimize polynomial and power models	Developing and optimizing a mathematical model to challenge an interpretation of a data set (Create your own data file)
15. Deeper Exploration of Logs and Exponentials	Applying calculus to the analysis and optimization of logarithmic and exponential models	Applying calculus skills to exponential functions in order to help a wine collector plan her wine storage for the future (Create your own data file)
16. Optimization in Several Variables	Defining constraints and performing constrained optimization using Excel's SOLVER routine	Determining optimal mix of advertising budget under uncertain conditions, using Solver (Data file)
17. Area Under the Curve	Evaluating definite integrals using both the Fundamental Theorem of Calculus and numerical methods to find the area under a curve.	Finding the area between curves to resolve a pricing dispute for a doll at Cool Toys for Tots (consumers' and producers' surplus). (Data file)

Dedication and Acknowledgements

First and foremost, this book is dedicated to Dr. Allen Emerson, my co-author and long-time friend, who passed away during the completion of this project. His hard work, tenacious intellect, and willingness try new ideas made this book possible. Our spouses also deserve a great deal of the credit for this work. Cheryl Forbes teaches writing and rhetoric, and her influence on Allen's approach to teaching mathematics was enormous. My wife, Brenda, has had a prfound influence on my approach to teaching overall and on helping me understand the business world enough to bring a new approach to mathematics into it. Both of them put up with our tendencies to lose sight of everything but this project, at times spending upwards of twelve straight hours a day trying to understand student learning in the course we wrote this book to support.

We would also like to thank Anne Geraci for her invaluable assistance. She has provided enormous editorial support in reviewing the materials and helping to prepare this updated edition of the textbook. Any errors, typos, or omissions are entirely due to our work and not her excellent reviewing of the material.

I would also like to thank Carol Freeman, the department of Mathematical and Computing Sciences at St. John Fisher College, and the School of Business at Fisher. They have provided us with opportunities to try new approaches to an old course and have supported our ideas, no matter how strange they seemed. The course we designed, and ultimately, the textbook we wrote, would also not have been possible without the assistance of many adjunct faculty members who helped us with suggestions, revisions and ideas: Mike Rotundo, Rebecca Tiffin, and Mary Ann Cape.

In addition, Ginger James provided us with invaluable assistance in the early years of the course, attending class, tutoring students, and offering suggestions while still an undergraduate at St. John Fisher College. We have also benefited from the able tutoring of several undergraduates, and thank all of them for their assistance in supporting the course.

Contents