Preface

In students' first course in Statistics, the focus was on learning about the process of conducting statistical investigations, most likely investigating research questions involving one variable or the association between two variables. This involved (1) identifying a research question, (2) gathering data to answer the research question, (3) examining the data, (4) testing hypotheses and estimating parameters using tests of significance and confidence intervals, (5) drawing conclusions about generalization and causation, and finally (6) reflecting on possible next steps in the research process.

This statistical investigation method mirrors the scientific method and is also utilized throughout this second course in Statistics. However, now the focus expands to include explicit consideration of both study design and statistical analysis strategies for investigating the simultaneous impact of two or more explanatory variables on a response variable of interest. Throughout this course, the focus will remain on the concept of *explaining variation in a response* variable of interest, and will emphasize a single, cohesive framework for thinking about explaining variation across study designs and variable types.

Approach

We adopt several distinctive features:

Keeping the focus on explaining variation. We introduce the theme of explained variation in the very first section and keep the focus on explained variation throughout the course. We do this by:

- Anticipating key themes of explained variation through intuitive examples and compelling visualizations beginning in the first section.
- Introducing and utilizing the **Sources of Variation** diagram as a natural visual representation of the components contributing to the variation in the response. This diagram also directly connects with notational representations of multivariable regression models, ANOVA tables which partition variation, and verbal explanations of relationships between variables. In particular, it focuses on (1) the variable of interest (response variable), (2) the sources of explained, and (3) unexplained variation in the response variable, and (4) the inclusion criteria and the design of the study.

Focusing on multivariable thinking. When developing students' conceptualization of multivariable thinking, it's easy to get bogged down in algebraic manipulations and algorithms. Instead, we try to develop students' multivariable thinking through

- **Visualizing adjusted vs. unadjusted associations by exploring what it means to subtract off effects and by thinking through the implications of choice of study design.** We use graphical displays to introduce students to the idea of whether and how much the relationship between the response variable and one explanatory variable can change in the presence of a second explanatory variable. These visuals will be used to emphasize the impact of study design (experimental vs. observational) on the estimates of effects of factors as well as the significance of factors.
- **Exploring patterns in the residuals and identifying how to explain more of the variation.** We will use any patterns leftover in residual plots to motivate the need to include additional explanatory variables in our variable relationship model so that we may be able to reduce the amount of unexplained variation.
- **Using simulation to understand study design, to consider test statistic choice, and to reinforce the core logic of inference.** Although we believe that simulation is possible to weave throughout the course in virtually every section, instead we choose to judiciously use simulation at key times where we believe simulation can maximize students' conceptual understanding of new concepts.

Integration of exposition, examples, and explorations. Every section includes one fully worked-out example that illustrates the ideas and methods presented, and at least one exploration that students work through to learn about and gain experience with applying the tools. We offer a great deal of flexibility for instructors to decide on the order in which they present these components, and what they will ask students to do in class vs. outside of class. *To facilitate this flexibility, examples and explorations within a section are written so that neither depends on the other,* allowing the instructor to present either one first.

Easy-to-use technology integrated throughout. Rather than ask students to start the course learning a statistical software package, we have designed easy-to-use web-applets that enable students to conduct all of the simulations and perform many of the analyses presented in this book. The emphasis of these applets is data visualization and conceptual understanding. Statistical software integration using R, JMP, or a package of the instructor's choice is presented carefully and in such a manner as to not distract from the main goal of conceptually understanding and modeling explained variation. Statistical software output from applets and packages such as R, Minitab, and JMP are incorporated in the examples, while the explorations give the instructor a choice of statistical software.

Tintle_FM.indd 8 09/07/20 4:46 PM

FM_Tintle1e_4PP

LUMINA DATAMATICS

Real data from genuine studies. We utilize real data from genuine research studies throughout the book using example-driven exposition to make the contexts more *meaningful*. These studies are taken from a variety of fields of application and popular culture. Each chapter also includes an end-of-chapter detailed investigation that uses data from genuine studies, including published research studies, giving students even more exposure to genuine applications of statistics.

Flexible content. This second course has been purposefully designed to follow nearly any algebra-based introductory statistics course (e.g., with or without simulation-based inference; AP statistics, etc.) so that it is accessible to audiences with varying backgrounds and experiences with mathematics and statistics. Many of the pedagogical and content approaches follow directly from *Introduction to Statistical Investigations*, and so our approach may be more familiar to students who have used that text before this one. Depending on first course content or how long it has been since the first course was taken, this review material may be focused on more/less/not at all.

Reinforcing key principles from the first course. Although we purposefully aim to introduce students to multivariable statistical thinking through the lens of explained variation, we also recognize that a second course in statistics must continue to reinforce key ideas from the first course in statistics. Thus, we have purposefully built in review questions and concepts throughout the book, to purposefully review key terms and learning objectives from the first course in the context of new learning objectives from this second course (e.g., overarching process of statistical inference, logic and scope of inference and connections to design and analysis strategies).

Changes in Content Sequencing and Coverage

A single framework for regression and ANOVA. Many second courses in statistics tend to more heavily emphasize regression over ANOVA or vice versa. Similarly, many second courses in statistics tend to emphasize experiments over observational studies or vice versa. We instead focus on illustrating a statistical model and tracking sources of variation as complementary and equivalent components of every analysis. This can provide different lenses through which variation can be explored and discussed, but which, ultimately, are equivalent and appropriate for both experiments and observational studies. Similarly, we aim to illustrate the key concepts of the course through a mix of both experiments and observational studies—highlighting the pros and cons of each but trying to provide students ample experience with both. Finally, this unified view of ANOVA and regression, complemented by our use of simulation/randomization techniques, allows us to take a simplifying view of validity conditions to help students focus primarily on the main concept of explained variation vs. becoming stuck in the weeds of validity condition nuances.

Descriptive statistics and data visualization. We take a case-study approach that focuses on the Statistical Investigation process as a whole. Thus, exploratory data analysis is integrated throughout this curriculum. The curriculum cycles through different types of data and numbers of variables in each chapter, so students are introduced to descriptive methods and multivariable data visualizations as they are necessary for each analysis.

Using simulation. We utilize simulation as a bridge to traditional asymptotic tests and concepts of multivariable thinking. While simulation methods are growing in importance in practice, our choice here is a pedagogical one—believing that judicious use of simulation deepens students' understanding of statistical concepts.

A balanced approach to data analysis. The six-step method of conducting a statistical investigation places data exploration before formal inference. Thus, we illustrate to students how data exploration takes place before formal inference, and sometimes formal inference is not needed. Emphasis is on what we can and cannot learn from available data, and how learning from data is an iterative process.

Changes in Pedagogy

Accessibility. Second courses in statistics frequently have mathematical prerequisites which limit their accessibility to a wide range of student audiences. We require only a single, algebra-based introductory statistics course as a prerequisite and include numerous opportunities for review and reinforcement of both first- and second-course concepts. Furthermore, we recognize that both accessibility and deep understanding benefit from a balanced approach to the use of words/verbal descriptions, data and conceptual visualizations, and mathematics equations/notations/symbols.

Active learning. In contrast to many other second courses, we utilize an active-learning approach, which immerses students into the statistical investigation method, and helps them engage with the various aspects of data collection and analysis. Each chapter contains a number of explorations for the students to complete, in addition to example-driven exposition of concepts. These materials allow for a variety of instructor-determined approaches to content delivery, including approaches where examples/concepts are presented first by the instructor and then explored by the student or vice versa.

Explorations. Student explorations involve a variety of learning experiences using computer-based simulations and visualizations, using Javascript applets, collecting data, running experiments, and interpreting results given by statistical software. The majority of explorations are flexibly designed

to be completed by students working individually, in small or large groups, either inside or outside of class.

Examples. Concepts are introduced using engaging examples explained in an easy-to-understand format that limits technical jargon and focuses on conceptual understanding. We have also included *Key Idea* and *Think About It* boxes to help students understand what they read, identify core concepts, and be engaged readers. Overall, we advocate utilizing a small amount of instructor-led interactive lecture and discussion, but mainly focusing on engaging and strengthening different student learning processes by way of a variety of active, self-discovery learning experiences for students.

Exercises and Investigations. Each chapter contains an extensive set of exercises. We also include an investigation at the end of each chapter: an in-depth exercise exploring the entire six-step statistical investigation method so that the single assignment can assess a variety of concepts. Each chapter also challenges students to develop their critical reading skills by including a research article for students to read followed by a series of questions about the article.

Real meaningful data. As we did in *Introduction to Statistical Investigations*, we continue to use real data *that matter* in our examples, explorations, investigations, and exercises. This is in keeping with the recently updated GAISE[1] recommendations that introductory statistics courses should "integrate real data with a context and a purpose." In our opinion, an introductory statistics course should be viewed less as a course in which students see "cute" but impractical illustrations of statistics in use, and more as a course that shows how statistics can be used to make decisions that have health, monetary, or other factors impacting hundreds, thousands, or millions of people. We find the benefits are two-fold: first in improving students' statistical literacy, and second in helping students to recognize that statistics is the indispensable, inter-disciplinary language of scientific research.

To the Student

We know from decades of experience as teachers of statistics that many students never master the most important, but hardest, ideas of our subject in their second course in statistics. Furthermore, we know that a great many students never even get the opportunity to take a second course in statistics because they need to take many additional mathematics courses first. These issues arise for a variety of reasons: (1) because the ideas are truly difficult; (2) because learning the formulas of statistics often gets in the way of learning the ideas of statistics and (3) because the hardest and most important ideas are too often saved for the end of the course, when time is running short.

This second course in statistics book is different. We show you the most important ideas up front, even though we know they are challenging. We downplay formulas, especially at the start, in order to put the ideas first; we also downplay mathematical details to focus more on conceptual understanding. This approach asks more of you up front, but we have become convinced from our own classes that, in the long run, this approach will pay off for you, the reader. Students leave our classes better prepared to use statistical thinking in their science, social science, and business courses, and in their careers after graduation.

At the same time, we also recognize that this approach may put you in an uncomfortable position. We are asking you at the beginning of the term to start working at understanding ideas that may take several weeks of thinking, effort, and practice to become clear. Many of the most important ideas in all subjects are like that. What we ask of you is continued effort and patience. In return, we offer our understanding that some of the goals we have set for you cannot be achieved in just a week or two or three.

[1]GAISE College Report ASA Revision Committee, "Guidelines for Assessment and Instruction in Statistics Education College Report 2016." Available at http://www.amstat.org/education/gaise.