

OpenIntro Statistics: an open-source textbook

1.0 INTRODUCTION

The traditional textbook, when well-written, can be a great thing. It has considerable breadth or depth of knowledge, and sometimes both. It is an excellent tool for self- or directed-study. A well written textbook offers stimulating examples and challenging problems for learning and assessment. It is nicely organized and well edited. Relatively unchanged in the past three or four centuries, the traditional textbook has long been the authoritative voice to which we turn when there isn't a knowledgeable human voice available. We are familiar with how to learn from them, and how to teach with them. Comedian Groucho Marx summarized the sentiment:

*Outside of a dog, a book is man's best friend.
Inside of a dog, it's too dark to read.*

In fact, there are just a few pervasive shortcomings of the well-written traditional textbook. Importantly, the traditional textbook can be rather exclusionary due to its price. Consider US community colleges, which serve approximately 13 million students each year. According to the American Association for Community Colleges, these students face average annual tuition and fees of approximately \$3,100. Furthermore, textbook prices average nearly \$1,100 per year for US college students ([American Association of Community Colleges 2013](#)). This creates a substantial barrier to students who wish to enter the education market as consumers. On the producers' side, publishing houses stand as gate-keepers that are bound by fiduciary responsibility to maximize profits when determining which books and which content to bring to press. In an increasingly digital era, the traditional textbook continues to be available almost exclusively in printed editions, reinforcing the stance that their knowledge transfer is proprietary. Electronic content that supplements the traditional textbook (or an electronic versions of the textbook) is becoming more common. However access to such resources generally requires students to purchase a temporary license from the publisher, which is often bundled as a mandatory expense with the textbook. Learners are expressly prohibited from modifying or owning the words they read or images they study. Indeed, the student is offered little opportunity to interact with the traditional textbook, besides the familiar habits of underlining and bending page corners.

Through these lenses the traditional textbook is seen as a highly successful teaching instrument, but one that has a handful of important shortcomings. Over the last decade publishers of the traditional textbook have been responding to these criticisms by devoting resources to developing electronic versions of their textbooks and making available supplementary materials such as web applets and online study materials. While some of these improvements have proven successful in supporting the changing landscape of student learning, other issues still remain to be resolved. In this paper we ask and answer a seemingly simple question: how can we keep what is good about textbooks while innovating where there are opportunities for improvement?

As the title of this paper suggests, we believe the key is to create an open-source textbook. Generally, *open source*, as a development model, promotes universal access via free license to a product's design or blueprint, and universal redistribution of that design or blueprint, including subsequent improvements to it by anyone (Gerber *et al.* 2010). In the context of a computer program, open source means that the source code is available to the general public for use and/or modification. For example, R is an open-source statistical programming language since anyone can download and modify the program itself, while SAS is proprietary since it costs money and its source cannot be edited by the licensee.

In Section 2, we describe how our content keeps pace with the latest advances in statistical pedagogy, while facilitating more traditional approaches to teaching introductory statistics. Section 2 also touches upon our style, which is modern but grounded. Section 3 focuses on our open-source approach, which represents a somewhat more substantial departure from tradition. As such, Section 3 is also where we discuss the important issue of credibility. We share our thoughts on the future of the textbook and provide other concluding remarks in Section 4.

2.0 CONTENT AND STYLE

OpenIntro takes cues on content and style from a range of sources, including established authorities like the American Statistical Association (ASA) and influential contemporary writers like Malcolm Gladwell, author of *Outliers* (Gladwell 2008), and brothers Chip and Dan Heath, authors of *Made to Stick* (Heath and Heath 2007). The goal is to create resources that are highly relevant and provide understanding and insights that remain with the students throughout their lives.

2.1 Content

Deciding upon specific content is a challenge. Introductory courses in statistics, the type

that require no preparation beyond algebra, are offered across a plethora of departments, including Mathematics, Statistics, Economics, Psychology, Public Health, Public Policy, Political Science, Environmental Science, and others. Furthermore, students may take their first course in statistics during high school, community college, university, or graduate school. Construct a matrix with these two dimensions, and you will find an active market for introductory statistics textbooks in many of the cells.

OpenIntro Statistics was written to provide necessary content to the widest range of students. As a result, *OpenIntro Statistics* overlaps with most general introductory courses in statistics.

The content of *OpenIntro Statistics* is most influenced by the ASA's Guidelines in Assessment and Instruction in Statistics Education (GAISE) reports and the Advanced Placement (AP) curriculum. Regarding content, the GAISE report calls for the use of real data, the utilization of technology for concepts and analysis, and assessment to improve and evaluate learning (Aliaga *et al.* 2010). *OpenIntro Statistics* uses data and software available in the accompanying `openintro` R package, which is a publicly-available package for instructors or others who may wish to access the textbook's data sets and custom functions used in generating the figures in the textbook (Diez *et al.* 2012). The textbook also includes over 550 exercises, many with full solutions for students to check their understanding. There is also an online quiz tool with a rapidly growing collection of online quiz questions to support the text.

OpenIntro Statistics balances newer statistical pedagogy focused on simulation and data analysis, with classic probability-driven approaches. The textbook is written so that instructors can seamlessly skip or rearrange chapters. For instance, linear regression (Chapter 7) can be visited following the first chapter, and probability (Chapter 2) is only a prerequisite in a few special topic sections. Certain sections, such as simulation and randomization, are key to offering instructors the opportunity to learn and teach modern methods. Many of these modern topics and their corresponding supplements, such as computer labs, are available at www.openintro.org and offer tools to make it easier for instructors to implement the GAISE technology recommendation.

Sections that can be skipped without affecting later material are clearly marked throughout the book as special topics, and the preface gives an overview of the modular nature of the content. However what sets *OpenIntro Statistics* apart from other textbooks with similar modular material is the ease of adopting only part of the textbook as it is relevant in the course. Instructors do not need to worry about justifying the cost of the book in order to assign only a few chapters or only use the book as supplementary material. For example, *OpenIntro Statistics* has been recommended as a reference book in a few massive online courses, e.g. *Statistics One* from Princeton University on Coursera (Andrew Conway,

September 2013).

2.2 Style

The stylistic approach of *OpenIntro Statistics* blends together ideas from contemporary writers and cues from the GAISE report. The influential book, *Made to Stick* (Heath and Heath 2007), extends concepts raised in *The Tipping Point* (Gladwell 2002), by suggesting stylistic steps that help people remember information being communicated. Interestingly, themes of these New York Times bestsellers overlap with the GAISE report.

OpenIntro Statistics focuses on presenting core ideas in simple and understandable terms that emphasize conceptual understanding. Each chapter introduces just a few principal ideas, repeatedly reinforces them, and draws connections to other fundamental concepts within the text. Furthermore, the textbook is built around concepts, not procedures.

OpenIntro Statistics emphasizes real data and examples from concrete and compelling stories throughout the text. Real and stimulating examples are key to engaging and motivating students. We also support teachers by providing open access to the data sets used in the textbook among others on our website, openintro.org. Statistical ideas are made concrete by introducing a specific case or example before providing a general explanation. These provide firm aspects that can be recognized, grasped, and referenced as students learn a new concept or technique.

With this “example first, method second” approach, each chapter of *OpenIntro Statistics* opens with a short paragraph on which examples will be presented, providing the students with concrete information on what to expect in the chapter. Introducing the examples first helps students learn to formulate good questions. Methodology is then presented as a way to use data to answer these questions appropriately. This approach addresses a common criticism of introductory statistics courses, which is that students have difficulty putting in perspective how statistical methods are applied to real life. In addition, using current and relevant examples contributes to the credibility of the textbook and makes it an appropriate text for statistical literacy courses.

3.0 AUTHOR, PUBLISHER, AND COMMUNITY

The traditional textbook is expected to be authoritative: trusted as being accurate, reliable, and clear. Publishing houses have historically signaled this authority by screening book proposals (and, implicitly, authors), as well as providing editorial assistance and coordinating peer reviews. As in other aspects of the book, *OpenIntro Statistics* draws heavily

from the traditional model, and then offers enhancements derived from being free and open source. Most importantly, being accessible and transparent empowers the community to play a leading role in validating the authority of *OpenIntro Statistics*. One of the ways by which we hope to further promote transparency in the future is via publishing a list of courses and faculty members that use *OpenIntro Statistics*. This aim of this list will be to implicitly connect users of the book and initiate dialogue about their experiences, which we believe will serve as a way to verify the quality of the book as well as to promote discussion on aspects of the book where there may be room for improvement.

3.1 Author

The authors of *OpenIntro Statistics* and authors of traditional textbooks are similar in many, but not all, ways. Our team has taught introductory statistics at about a half dozen schools, ranging from land-grant technical universities to the Ivy League. We studied under textbook authors during our doctoral and postdoctoral training. Additionally, though relatively young, we are active in pedagogical research and course design. While there are many similarities, it is easy to identify at least one *prime facie* difference between the authors of *OpenIntro Statistics* and authors of other introductory statistics books: both groups may have similar motives for wishing to write a textbook, but the authors of the traditional textbook must then be selected by a publisher based on their potential to maximize profits.

Authorial tone is also important when writing a textbook. The tone in *OpenIntro Statistics* is professional and empowering. We endeavor to communicate information to the students in a professional manner, consistently focusing on the core material and omitting gratuitous photos and content. Furthermore, instead of attempting humor, we pursue students' attention via current and interesting examples, research questions, and data sets. To avoid oversimplifying material, which may leave unanswered questions and disrupt credibility, *OpenIntro Statistics* seeks to create a gradual slope up the intellectual path. For example, we introduce easier introductory questions within the text. When a student gets stuck, we support them by providing full solutions to nearly all of in-text exercises in footnotes. We communicate to the student: we think you can do it, and if you struggle, we can help. Our hope is that this approach leaves the students feeling empowered.

3.2 Publisher

While publishing companies continue to be prominent players in the textbook market, their necessity is diminishing. Traditional tasks such as vetting authors, editing drafts, printing, and marketing finished textbooks are taking on new forms. These tasks are also being

organized and performed by new parties. As we describe in the next section, the model for *OpenIntro Statistics* empowers the community to fill many of the roles exclusive the traditional publisher.

First, however, it is useful to highlight some similarities between the publishing cycle of *OpenIntro Statistics* and the traditional textbook, especially in regards to editing and printing. The editorial process for *OpenIntro Statistics* included extensive review, with one author playing a dedicated role as internal editor. Then, as is customary in the traditional framework, portions of the book were reviewed and edited by colleagues in the education community. Printing of *OpenIntro Statistics* is performed by the on-demand printer CreateSpace. The resulting paperbacks are available on amazon.com, just like other textbooks, but at a fraction of the cost: under \$10.

It is worth noting a few differences that result from circumventing the traditional publishing firm, and the impacts these have on cost. Perhaps the most significant topic in this regard is marketing. The amount of resources that a publishing company devotes to marketing a textbook is truly humbling for our volunteer-driven organization. Traditional textbooks certainly have a head start in getting across the desks of potential adopters quickly. However, like other peripheral expenses, the costs associated with marketing become part of the substantial price tag.

3.3 Community

At the college level and beyond, instructors bear the responsibility of selecting the best book for their students. This can be a substantial challenge. Furthermore, it is often impractical to solicit feedback from students until after the students and the instructor have already committed to using a particular book for the duration of a course. With prices of the traditional textbook often topping \$150, it is generally unrealistic for students to procure and comment on alternative options.

The OpenIntro model empowers experimentation, customization, and evaluation of the textbook without the barrier of a substantial purchase price. This allows for the gradual review of *OpenIntro Statistics* over a period, perhaps concurrently with comparison to other texts, to allow for enhanced evaluation and smoother transitions. A free textbook can easily be added as a course supplement. At the end of a course, students can easily be surveyed for their impressions. Indeed, we've already received favorable feedback from a handful of instructors who used *OpenIntro Statistics* as a supplemental text.

4.0 THE FUTURE

As technology rapidly evolves, we are focused on balancing customization and standardization. How do we develop resources that are flexible enough to meet the diverse learning styles and needs of students around the world while still producing tools that fit together in a unified framework?

We should be mindful of diversity in the way students learn. Recent reports from a pilot study for e-textbooks at Indiana University indicated that over a quarter of participating students believed they would have learned more using a paper textbook than the e-textbook provided in the course (Dennis *et al.* 2010). This is a good example of how different students learn in different ways and require different resources. Rather than focusing exclusively on what format works best for the largest proportion of students, we should cater to multiple styles of learning, from electronic textbooks to paper textbooks to online videos and more interactive learning experiences.

Students have diverse needs that call for a range of resources that complement and support each other in a unified framework. In addition, we have observed that even an individual student will use the textbook in multiple ways. Classroom testing of *OpenIntro Statistics* has revealed that, given the opportunity, students will make use of the textbook in its different forms (online, self-printed, and purchased) in different circumstances. Making the text available in a variety of formats enables the students to move easily between formats depending on what is convenient, and helps us reach a wider range of learning needs and preferences.

Traditional publishers also seem to agree that providing a range of resources along with a textbook is necessary. It is becoming increasingly common for textbooks to be bundled with online supplements. Unlike publishers who seek to develop proprietary and exclusive online tools, we believe it is important for online resources to be free and community driven.

Another quality that sets open-source books apart from the traditional textbook is the flexibility to update material, such as current examples, in-the-news articles, etc. Since the preliminary edition of *OpenIntro Statistics* which was published in 2010, we have made two major and one minor revision to the textbook. The majority of these revisions were in the examples and exercises presented in the book, and with each revision we incorporated more up-to-date data sets. With an open-source text, it is also trivial to revise typos directly in the book as opposed to having to provide a separate errata that most students tend to overlook. The ease of making revisions and updates to keep the text fresh and error-free without having to justify the cost of changing editions to students (and the ability of an instructor to continue using an earlier edition, if that is desired) is an important quality that sets apart the open-source textbook from its traditional counterparts.

While retaining many of the fundamentals of the traditional textbook, *OpenIntro Statistics* is both free and comes with an ever-growing set of free, unified resources. The traditional publisher has offered many helpful services, but we believe technological innovation has facilitated an era in which the community can be leveraged to produce high quality educational materials without a traditional publisher. Our model promotes transparency, teaching tool diversity, and lowers barriers to education and learning.

5.0 REFERENCES

- Aliaga M, Cobb G, Cuff C, Garfield J, Gould R, Lock R, Moore T, Rossman A, Stephenson B, Utts J, Velleman P, Witmer J (2010). “GAISE College Report.” *American Statistical Association website*. URL <http://www.amstat.org/education/gaise>.
- American Association of Community Colleges (2013). “2013 Community College Fact Sheet.”
- Dennis A, Duffy T, Morrone A (2010). “Indiana University e-textbooks Project.”
- Diez DM, Barr CD, Çetinkaya-Rundel M (2012). *openintro: OpenIntro data sets and supplement functions*. R package version 1.4, URL <http://cran.r-project.org/web/packages/openintro/index.html>.
- Gerber A, Molefe O, van der Merwe A (2010). “Documenting open source migration processes for re-use.” In “Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists,” pp. 75–85. ACM.
- Gladwell M (2002). *The Tipping Point: How Little Things Can Make a Big Difference*. Back Bay Books.
- Gladwell M (2008). *Outliers: The Story of Success*. Back Bay Books.
- Heath C, Heath D (2007). *Made to Stick: Why Some Ideas Survive and Others Die*. Random House.