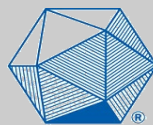


November
30, 2019

StatPREP



MAA

MATHEMATICAL ASSOCIATION OF AMERICA

NEWSLETTER

STATISTICAL INFERENCE VIA DATA SCIENCE: A MODERNDIVE INTO R AND THE TIDYVERSE **BOOK REVIEW**

BY DONNA LALONDE

"Finally, feel free to modify the book as you wish for your own needs, but please list the authors at the top of index.Rmd as 'Chester Ismay, Albert Y. Kim, and YOU!'" This quote is from the about section of the book *Statistical Inference via Data Science*. So, this book is not only a resource but an opportunity to be an author! The complete title of this book is *Statistical Inference via Data Science: A moderndive into R and the tidyverse*. The authors credit *Mathematical Statistics with Resampling and R* by Chihara and Hesterberg, *OpenIntro: Intro Stat with Randomization and Simulation* by Diez, Barr, and Çetinkaya-Rundel, and *R for Data Science* by Golemund and Wickham for inspiration for this book which is available online at <https://moderndive.com/>.

This book was designed to support instructors of Introductory Statistics courses and it is written for students with no prior algebra, calculus or programming experience. principle. The three parts of the book - data science via the tidyverse, data modeling, and statistical inference - are covered in 11 chapters. The chapters include data visualization, data wrangling, basic regression, multiple regression, sampling, bootstrapping and confidence intervals, hypothesis testing, inference with regression, and telling the story with data. The chapters have "learning checks" for which the solutions are included as an appendix. Each chapter begins with a list and brief description of the required R packages.

....continued on page 2

WHO'S WHO:

LEADERSHIP TEAM

Mike Brilleslyper,
Air Force Academy

Jenna Carpenter,
Campbell University

Danny Kaplan,
Macalester College

Kathryn Kozak
Coconino Community
College

Donna LaLonde,
ASA

Ambika Silva
College of the Canyons

Rachel Levy
MAA

HUB LEADERS

Joe Roith, St. Olaf's Col-
lege, Northfield, MN (2017-
18)

Ambika Silva, College of the
Canyons, Santa Clarita, CA
(2017-18)

Helen Burn, Highline Col-
lege, Seattle, WA (2018-19)

Hwayeon Ryu, Elon Univer-
sity, Elon, NC (2018-19)

Carol Howald, Howard Com-
munity College, Columbia,
MD (2019-2020)

Thomas Kinzeler, Tarrant
County College, Fort Worth,
TX (2019-2010)

Rona Axelrod, Florida SW
State College, Fort Myers,
FL (2020-2021)

Brooke Orosz, Essex Coun-
ty College, Newark, NJ
(2020-2021)

Support for this MAA Program is provided by NSF DUE-1626337



STATISTICAL INFERENCE VIA DATA SCIENCE REVIEW continued..

The topics are motivated by case studies or scenarios. For example, the discussion on basic regression is framed using a research on the factors that influence course evaluations of instructors, and the sampling discussion utilizes a 2013 poll conducted by National Public Radio. The scenarios and case studies could be used to augment a primary text in a course.

The final chapter is titled, "Tell the Story with Data." This chapter reinforces the need for good communication skills. The case study used in this chapter deals with housing costs in Seattle. For instructors interested in adding projects and presentations to their courses, this chapter would be a useful resource. It is a strength that most of the chapters in this book can be utilized as stand alone resources.

Dive into *Statistical Inference via Data Science: A modern dive into R and the tidyverse* as you are planning for the spring semester!



NEW RESOURCES FOR TEACHING STATISTICS ON MAA CONNECT

BY KATE KOZAK

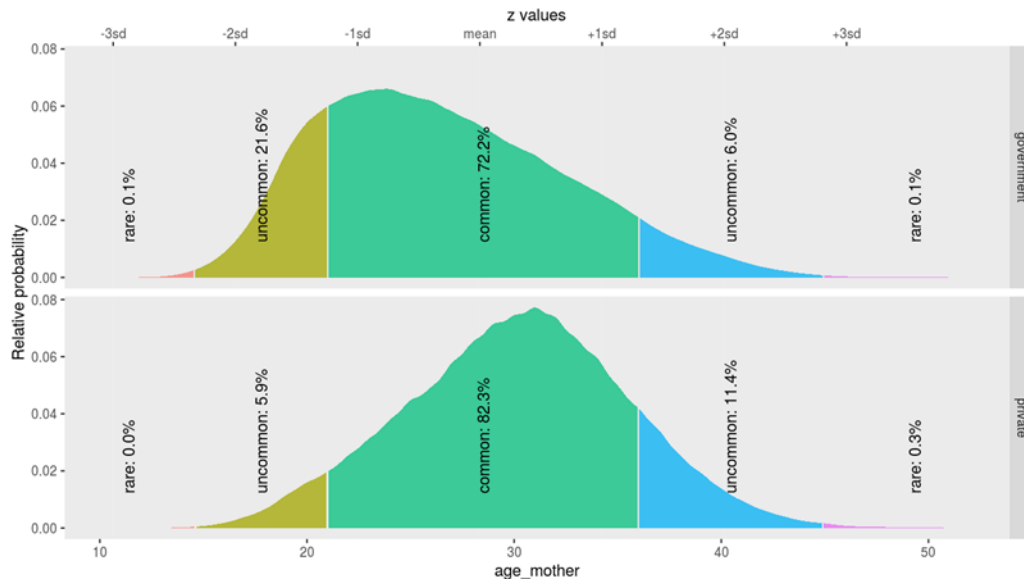
On MAA Connect, there is a post called, "Another resource for teaching introductory statistics." This resource highlights new apps called "Happy Apps," created by Dan Adrian, an assistant professor of statistics at Grand Valley State University in Grand Rapids, MI. How do the Happy Apps relate to the StatPREP project's Little Apps? The Little Apps are rooted in displays of data; statistical concepts are always directly put in the context of such displays. The Happy Apps have graphics oriented to methods and the statistical theory behind them. These are two different pedagogies for teaching statistics.

Go to MAA Connect to access the post that has links to access the "Happy Apps."

DISCOVERING STATPREP.ORG - ACTIVITIES FOR LITTLE APPS

BY AMBIKA SILVA

Have you checked out the “Activities for Little Apps” page? You can find this under the “Resources” tab on the main page of the StatPREP website, or by going to <http://statprep.org/resources/>. This is a collection of activities for you to use in your classes. The web pages give you insights on how to use the little apps with your students. There are word documents that you can use with your students to help them gain a deeper understanding of statistical concepts and use data sets that help to explore these concepts. You’re able to change them to how you see fit for your classroom. For instance, if you go down to the “Rare and Common” page, you will see a great graphic that shows what common, uncommon, and rare may look like for the age of a mother giving birth.



These are graphics from the Common and Rare Values Little App . You can let your students play with the applet by adjusting what they feel should be common and rare. Need something more structured? At the bottom of the “Rare and Common” page you will find a link to one or more activities associated with the Little App.

....continued on page 4

DISCOVERING STATPREP.ORG Continued...

Suggested StatPREP 101 activities

- Common, uncommon, and rare

Clicking on the hyperlink under “Suggested StatPREP 101 Activities” will bring you to a page that goes over an activity made by someone on the StatPREP team. It will also give you a chance to open it up as a word document.

COMMON, UNCOMMON, AND RARE

Get formatted versions: [Word](#) : [PDF](#)

Orientation

Change it, modify it, give more room in places you want your students to write more. If you do, we would love to hear about how it worked at MAA Connect! (<https://connect.maa.org/home>)



Not on MAA Connect? Learn how to join in the fun by watching our recent October webinar, with host Deveney Brown, MAA Communities Coordinator.

<http://statprep.org/webinar-series/>

NATIONAL CONFERENCES

BY KATE KOZAK

Presentations about statistics and data science can be found at many national conferences. As an example, the American Mathematical Society (AMS) Mini-Conference on Education was held in October in Washington, DC. The entire mini-conference was dedicated to the topic of data science. Another example is the American Mathematical Association of Two-Year Colleges (AMATYC) Annual Conference in Milwaukee, WI, which took place recently. At that conference there were numerous sessions on both statistics and data science. In addition, AMATYC has a statistics committee and a data science subcommittee. You can join these committees to learn more about these topics and participate in the discussions about statistics and data science education.

At the Joint Mathematics Meetings (JMM) Conference in Denver, CO, in January 2020, there will be a number of sessions devoted to statistics, including a few mini courses on statistics education. At the National Organization for Student Success (NOSS, formally NADE) conference in Nashville, TN, in March there will be a few sessions on statistics. Then there will be the NCTM Centennial & Exposition Conference in Chicago in April. The importance of statistics in K-12 should mean that there will be many sessions on statistics there. NCTM will be making a change to hold their conferences in the fall starting with their October 2020 conference. In August, there is MAA's conference called MathFest. This year MathFest will be held in Philadelphia, PA, at the end of July, and then the Joint Statistics Meetings (JSM) will be held in Philadelphia right after MathFest. As always, MathFest will have many statistical sessions, and of course JSM will focus on statistics. Then, AMATYC will host its annual conference in Spokane during November 2020.

Hopefully there will be one or more national conferences near you that you can attend to learn more about statistics education. It is hoped that your StatPREP exposure to data-centric statistics will encourage you to learn more about statistics and how to effectively teach statistics. If you are feeling adventurous, you may even consider presenting at one of these conferences! All of them have their own deadline for submitting a proposal. Most are a number of months before the conference itself. As an example, the AMATYC Annual Conference in Spokane has a deadline of February 1, 2020, for submitting a proposal to present.

....continued on page 6

NATIONAL CONFERENCES Continued...

Another opportunity for professional development this year is the International Congress of Mathematics Education (ICME). ICME only occurs every four years, and the next congress is in Shanghai, China, July 12 to July 19. This is a great opportunity to learn how mathematics is taught around the world. There will be a focus on statistics at this congress.

No matter where you go for professional development, remember a national conference is a great way to learn new and exciting topics from people around the nation. Learning from others helps you grow as a faculty member, and we all benefit from this professional growth throughout our careers.

Want more StatPREP? Check out:
<http://statprep.org/>

REVIEW OF *THE ART OF STATISTICS: LEARNING FROM DATA*

BY DANIEL KAPLAN

The Art of Statistics: Learning from Data by David Spiegelhalter (2019) Pelican Books 978-00241-39863-0

David Spiegelhalter is an eminent British statistician and former president of the Royal Statistical Society. He is well known for - among other things - his work in forensic statistics in high-profile cases and as a public expositor of statistics.

So when I heard about his new book, *The Art of Statistics*, I anticipated a rich collection of case studies drawn from his own experience and the experiences of the many statisticians he is in contact with. And this is a "trade book" (in the publishing lingo), which means it aims at a general audience. As it happens, most of our students are also drawn from a general audience, so I anticipated being able to find compelling and current classroom examples of the power of statistics to inform diverse situations.

There are many such examples in the book: detecting a murderous physician, cancer and bacon consumption, statins and heart disease, identifying the skeleton of Richard III, and others. Fun!

...continued on page 7

REVIEW OF *THE ART OF STATISTICS*

Continued...

But *The Art of Statistics* is not a catalog of professional anecdotes. It's a kind of textbook. I don't mean in the superficial sense of heavy, expensive, and presented in 4-color printing with distracting digression boxes. It has no exercises. I mean that the chapters sequentially lay out the foundations of statistical reasoning and offer a broad view of the whole field.

It's not the kind of statistical textbook you're used to. It's not the traditional "bag of tools," most of which date from around 1900. It's written with the current era in mind, where statistics is a team player rather than the referee for journal publication:

"We are in an age of data science, in which large and complex data sets are collected from routine sources such as traffic monitors, social media posts, and internet purchases, and used as a basis for technological innovations such as optimizing travel routes, targeted advertising, or purchase recommendation systems.... Statistical training is increasingly seen as just one necessary component of being a data scientist, together with skills in data management, programming and algorithm development, as well as proper knowledge of the subject matter."

The chapters are short and say what needs to be said concisely. For instance, risk is a ratio. The elaborations are about communication, e.g. how risk is presented, interpreted, and often mis-interpreted. Means are an average. Elaborations: looking at data is more informative than averages; our interest may be in something other than the average. And what is it that we're looking for when we use data?

By chapter 4, we've made a sharp turn from textbook convention. The title is "What Causes What?" Spiegelhalter thinks that causation is a statistical topic, to be taken seriously rather than dismissed with a simple mantra.

Chapter 5 introduces multiple regression, an important tool for making responsible statements about causation. It's the "multiple" that makes regression a tool for investigating causation; the "multiple" provides the means to incorporate what we know about mechanisms and how to mitigate the problems of confounding.

....continued on page 8



REVIEW OF *THE ART OF STATISTICS* Continued...

Chapter 6, "Algorithms, analytics, and prediction," highlights the many techniques there are for making a prediction, with the central but subtle question being: How good are the predictions?

There's no real mention of probability until mid-way in the book. There the emphasis is on communicating. A section is tellingly entitled, "The rules of probability made, possibly, a bit simpler." He builds on "a more intuitive idea that has been shown in many psychology experiments to improve people's reasoning about probability."

For educators, especially for mathematicians teaching statistics, it's worth quoting Spiegelhalter on the (relative lack of) importance of probability to statistics:

"Traditionally, a statistics course would start with probability ... but this rather mathematical initiation can be an obstruction to grasping all the important ideas in the preceding chapters that did not require probability theory. In contrast, this book is part of what could be called a new wave in statistics teaching, in which formal probability theory as a basis for statistical inference does not come in till much later."

When this is done in chapter 9, "Putting probability and statistics together," there is a

warning: "This is perhaps the most challenging chapter in this book." Think about that: he's saying that statistical inference is more challenging than multiple regression and conditional probability.

And what do we gain from focusing on this challenging topic. Rigor? Spiegelhalter quotes Jerome Cornfield (who took on Ronald Fisher about mathematical inference on smoking and cancer, and won!): "The paradox is that a solid structure of permanent value [specifically, the textbook notion of hypothesis testing which is a strange amalgam of the conflicting Fisherian and Neyman-Pearson perspectives] has, nevertheless, emerged, lacking only the firm logical foundation on which it was originally thought to have been built." Exactitude? Spiegelhalter points out that experience shows confidence intervals calculated using statistical theory systematically and substantially understate uncertainty. The next time you teach about calculating margins of error, and use 1.96 as a multiplier rather than 2 or emphasize that the z-distribution is different from t, perhaps you should share Spiegelhalter's heuristic: "[A]ny quoted margin of error in a poll should be doubled to allow for systematic errors made in the polling."

The Art of Statistics wraps up with two chapters whose titles say a lot: "How things go

...continued on page 9



REVIEW OF *THE ART OF STATISTICS*

Continued...

wrong” about, in part, the crisis of replicability in science; “How we can do better,” which offers ten questions to ask when “confronted with a claim based on statistical evidence.”

Almost 60 years ago, the renowned statistician John Tukey proposed a field of “data analysis,” which he saw as a needed course correction from the overly theoretical emphasis of mathematical statistics. Some of his innovations - stem-and-leaf plots, box-and-whisker plots - still appear in introductory statistics, but he had a broader program in mind which, today, would be called “data science.” He proposed that one of the essential components of “data analysis” as a field be the study of data analysis itself: its successes and failures, techniques that work and techniques that don’t. He wrote, in 1961, that “we need to give up the vain hope that data analysis can be founded upon a logico-deductive system like Euclidean plane geometry (or some form of the propositional calculus) and to face up to the fact that data analysis is intrinsically an empirical science.” The Art of Statistics is the result of thinking about statistics in this way: what works, what doesn’t, what can be effectively communicated and what can’t, and how to avoid unnecessarily centering statistics on nice theoretical notions that are relevant to only a narrow slice of the challenges faced when trying to extract information from data.

SAVE THE DATE!

Summer 2020 StatPREP Workshops

Fort Myers, FL: May 29 - 30, 2020, Florida Southwestern State College

Fort Worth, TX: June 3 - 4, 2020, Tarrant County College

New Jersey: June 5 - 6, 2020, Essex County College

Columbia, MD: June 12-13, 2020, Howard Community College



UPCOMING EVENTS

DECEMBER 2019

SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6	7
8	9	10 Consulting Office Hours Webinar	11	12	13	14
15	16	17	18	19	20 Consulting Office Hours	21
22	23	24	25	26	27	28
29	30	31				

DECEMBER WEBINAR

THE COMPACT GUIDE TO CLASSICAL INFERENCE: WHY?

In December 2019, StatPREP.org starts the serial publication of a short and highly specialized book: The Compact Guide to Classical Inference. Keeping in mind that StatPREP is about centering statistics education on real data, this book is an anomaly: the introduction of a reframed mathematical approach to statistical inference. The reframed approach unifies into a simple, single standard procedure all the inferential settings found in intro stats and carries forward naturally to apply to "advanced" settings such as multiple regression. The webinar will be a conversation between Donna and Danny about how the many formulae and probability tables of historical statistics can be unified into a single simple test statistic, F , and an expression for the confidence interval in terms of F .

Hosted by Donna LaLonde & Danny Kaplan



MAA
MATHEMATICAL ASSOCIATION OF AMERICA



ASA