



Article from

Predictive Analytics & Futurism

December 2018

Issue 19

Book Review: Actuarial Statistics With R

By Mary Pat Campbell

BOOK DETAILS:

Actuarial Statistics with R: Theory and Case Studies
Guohun Gan, PhD, FSA; Emiliano A. Valdez, PhD, FSA
Publisher: ACTEX Learning
Publication date: 2018

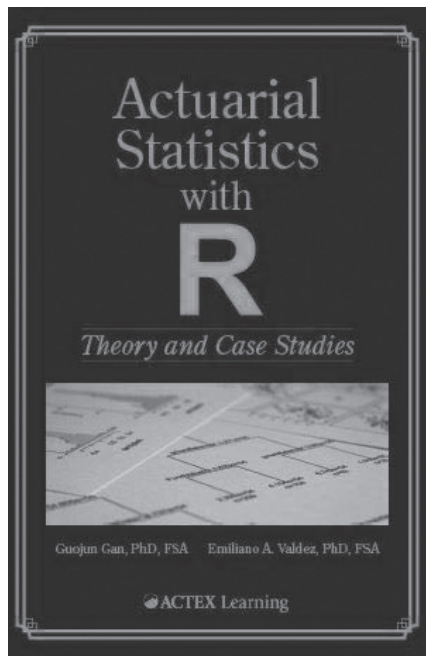
With the hotness of data science and predictive analytics in the actuarial world, many actuaries are looking to get up to speed on the latest approaches. Indeed, I wrote an article in December 2015 about getting started in predictive analytics, recommending various books and online courses.

However, from an actuarial viewpoint, the main problem was that generally one was using non-actuarial examples of data and models to work on. When one is a working actuary, it would be better if one could learn the material and applications specific to actuarial work at the same time.

With the hotness of data science and predictive analytics in the actuarial world, many actuaries are looking to get up to speed ...

This textbook, *Actuarial Statistics with R*, helps to bridge this gap. I received a review copy of this new text from ACTEX Learning, along with a copy of its solutions manual. This does look as if it's developed originally with college students in mind. According to the publisher:

“The content covers several topics on data analytics that have been prescribed by the International Actuarial Association. In particular, it has been designed to cover the learning objectives for the Statistics for Risk Modeling (SRM) Exam established by the Society of Actuaries. Some materials from this textbook also cover parts of



the syllabus for the Modern Actuarial Statistics (MAS-I and MAS-II) Exams of the Casualty Actuarial Society.”¹

While this text is partly actuarial exam related, I'm reviewing this from the point of view of a practitioner wanting to learn some statistical techniques, with a focus on using it for actuarial purposes.

INTRODUCTION TO ACTUARIAL STATISTICAL TECHNIQUES

The book is structured into four large sections: supervised learning, unsupervised learning, time series models and simulation. Within each section, the introduction of the specific technique and its underlying structure or theory is given, followed immediately by one or more case study chapters. For example, the chapter on generalized linear models is followed by three case study chapters: predicting demand for term life insurance, modeling number of auto claims, and modeling the loss severity of auto claims.

I find this case study approach to be very effective. While for college students, these case studies may be introducing unfamiliar insurance concepts, for the practicing actuary, it provides a base of familiarity. In some of these cases, as with regression, many of us had done these sorts of fits before if we've done any modeling, but perhaps we were not so familiar with using R.

Within chapters themselves, there is an alternation between mathematical notation and qualitative explanation of techniques with simple code examples implementing various

calculations. The code and data used in each chapter are also downloadable from the publisher's website.² It's freely available to the public, though without the context of the book itself, you may get a little confused. There are some comments in the code, but most of the explanation is within the text itself.

The level of mathematical knowledge needed is college-level calculus and linear algebra—there are summation formulas galore, and matrix multiplication makes its appearance. Many of the exercises in the text are more set up for college students or those preparing for the relevant actuarial exams than for the practitioner, but I found the specific symbolic derivations to be important in showing connections between various concepts, such as the connection between the leverage of a point in a linear regression (a measure of how far a point is from other data points, based on the independent variables) and the variance of the residual for a point at that independent variable.

CAVEATS FOR BEGINNERS: DEALING WITH R AND OTHER ISSUES

As the title says, the specific case studies are worked-through in R, and even when introducing the theory, they show R code that calculates the quantities just given via mathematical notation.

The problem starts if you are completely new to R. There is an appendix to provide some introduction, and it is about 50 pages long, so it's not all that bare bones. However, there can be practical issues for someone going to download R³ and then starting to play with it.

The level of introduction given is still very basic. You are shown foundational data structures and codes, but nothing is particularly complicated. The textbook in terms of technique and code is an introductory level—if you want to start making changes beyond what is given, you will likely need additional resources.

If you are a complete beginner to R, you may wish to look at the R Project's "Introduction to R."⁴ From the authors' own listed references, you may wish to look into Albert and Rizzo's *R by Example*⁵, which also takes an approach of going through more basic statistical analyses as a way to teach both the language and basic statistics at the same time.

The caveats also pertain to the content in the book in general. This is an introduction to these techniques, providing information on model evaluation and some practical approaches to improving models. There are more advanced texts and materials available once one has learned these basics. It helps to have some experience trying out the basic techniques before trying full-blown approaches.

I recommend those using this text to actually type the code in as one follows along in the text. As mentioned earlier, all the code can be downloaded from the publisher's site, but I think it helps to type in the commands on one's own to think through what is going on. Copying and pasting, or even just running the files as R scripts, means that you may miss some important details.

Finally, a note about some details that I really liked: I liked how the code was presented in the text. As is usual in texts with code, there is a different font from the standard text (specifically a typewriter-looking fixed-width type instead of the more aesthetically pleasing variable-width types). In addition to that, code is set apart in bordered regions, and, most importantly, the lines are numbered.

Given the way the text is set it can be slightly difficult to tell the code from the calculation result (the difference is whether the line starts with > or not), but the lightly-colored line numbers to the left of the code box still helps get one a handle on code blocks that can be as long as 42 lines between code and results.

In addition to a good presentation of the code, there are useful indices in the back—the usual type of index by topic, but also an index of R functions and symbols.

The authors pack quite a bit of information in this relatively short book (374 pages, including indices and excluding front matter), and I think this would be a handy introduction to any working actuary wanting to get started in statistical modeling. ■



Mary Pat Campbell, FSA, MAAA, PRM, is VP, insurance research at Conning in Hartford, Conn. She can be reached at marypat.campbell@gmail.com.

ENDNOTE

- 1 Gan, Guojun and Valdez, Emiliano A. Actuarial Statistics with R: Theory and Case Studies, page xvii
- 2 Link to files are here: https://www.actexamdriver.com/Assets/ClientDocs/supplement/Supplement_Actuarial%20Statistics%20with%20R.zip (accessed September 11, 2018). If the link moves, the files will be found via the Product Supplements section of the publisher's website (<https://www.actexamdriver.com/>)
- 3 R is downloadable at the R Project for Statistical Computing: <https://www.r-project.org/>
- 4 <https://cran.r-project.org/doc/manuals/r-release/R-intro.html> Accessed September 11, 2018
- 5 Albert, J. and Rizzo, M. (2001). R by Example. Springer, New York, NY. <https://amzn.to/2Qn4uzZ>