Analytic Methods in Sports

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Analytic Methods in Sports

Using Mathematics and Statistics to Understand Data from Baseball, Football, Basketball, and Other Sports

Second Edition

Thomas A. Severini
Northwestern University

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper

International Standard Book Number-13: 978-0-367-25207-6 (Hardback)
978-0-367-46938-2 (Paperback)

Library of Congress Cataloging-in-Publication Data

Names: Severini, Thomas A. (Thomas Alan), 1959- author.
Title: Analytic methods in sports : using mathematics and statistics to understand data from baseball, football, basketball, and other sports / Thomas A. Severini.
Description: 2nd edition. | Boca Raton, Florida : CRC Press, 2020. | Includes bibliographical references and index. | Summary: "Analytic Methods in Sports: Using Mathematics and Statistics to Understand Data from Baseball, Football, Basketball, and Other Sports, 2nd Edition provides a concise yet thorough introduction to the analytic and statistical methods that are useful in studying sports. It explains how to apply the methods to sports data and interpret the results, demonstrating that the analysis of sports data is often different from standard statistical analyses. The book integrates a large number of motivating sports examples throughout and offers guidance on computation and suggestions for further reading in each chapter"--Provided by publisher.
Identifiers: LCCN 2019056013 (print) | LCCN 2019056014 (ebook) | ISBN 9780367252076 (hardback) | ISBN 9780367252090 (ebook)
Subjects: LCSH: Sports--Data processing. | Sports--Mathematical models. | Sports--Statistical methods.
Classification: LCC GV568 .S48 2020 (print) | LCC GV568 (ebook) | DDC 796.02/1--dc23
LC record available at https://lccn.loc.gov/2019056013
LC ebook record available at https://lccn.loc.gov/2019056014

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To Kate, Tony, Joe, and Lisa

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contents

Preface

As expected, in the five years since the first edition of this book was published the use of mathematical and statistical methods in sports has continued to grow. Methodology that was once reserved for those sports fans with a background and interest in mathematics now influences nearly all sports analyses. My goal in the second edition is to better reflect this current state of analytic methods in sports.

Specifically, there are three important additions to the second edition. In the first edition, all calculations are performed using Excel, because of its wide availability and ease of use. However, the use of R for sports analyses has greatly increased over the past few years. Hence, the second edition uses both Excel and R. This will be more convenient for readers who are familiar with R and, for those who are not, this material can serve as an introduction to the use of R.

The first edition focuses on those core topics in probability and statistics that play a central role when analyzing and understanding sports data. However, as sports analysts become more familiar with statistical methodology, they naturally turn to more sophisticated methods. Thus, a chapter discussing several more advanced methods has been added, covering topics such as regression models for a binary response, the use of random effects to model and understand variation, multilevel models, spline methods, and principal components analysis, among others. Because of the specialized software needed for such methods, this chapter uses R exclusively.

Finally, with interest in the use of analytic methods in sports at an all-time high, many colleges offer courses in sports statistics. To enhance the value of this book for such courses, as well as to aid those engaged in independent study, exercises have been added to the end of each chapter.

I would like to thank a number of readers who sent comments and corrections on the first edition, as well as reviewers who provided recommendations for the second edition. I would also like to thank Rob Calver for continued useful advice and Karla Engel for her help in preparing the manuscript.

Thomas A. Severini
severini@northwestern.edu
Evanston, Illinois

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Preface to the First Edition

One of the greatest changes in the sports world over the past 20 years or so has been the use of mathematical methods, together with the vast amounts of data now available, to analyze performances, recognize trends and patterns, and predict results. Starting with the "sabermetricians" analyzing baseball statistics, both amateur and professional statisticians have contributed to our knowledge of a wide range of sports.

There is no shortage of books, articles, and blogs that analyze data, develop new statistics, and provide insights on athletes and teams. Although there are exceptions (such as *The Book*, by Tango, Lichtman, and Dolphin [2007] and much of the work presented on FanGraphs at http://www.fangraphs.com), in many cases the analyses are ad hoc, developed by those with a knack for numbers and a love of sports. The fact that their methods have worked so well, and their results have been so useful, is a testament to the skill and ingenuity of the analysts.

However, just as formal methods of statistical analysis have contributed to, and improved, nearly every field of science and social science, these methods can also improve the analysis of sports data. Even when such methods are used in the context of sports, rarely is enough detail given for the beginner to understand the methodology. Although such methods are routinely covered in university courses on statistics, a student might need to take three or more courses before techniques that are useful in analyzing sports data are covered. Furthermore, the analysis of sports data has special features that generally are not covered in standard statistics courses and texts.

The goal of this book is to provide a concise but thorough introduction to the analytic and statistical methods that are useful in studying sports. It focuses on the application of the methods to sports data and the interpretation of the results; the book is designed for readers who are comfortable with mathematics but who have no previous background in statistics.

It is sometimes said that the key to an interesting and useful analysis in sports is asking the right questions. That may be true, but it is also important to know how to answer the questions. In this book, I try to give the reader the tools needed to answer questions of interest.

I would like to thank Karla Engel, who suggested the project and who helped in preparing the manuscript, including making many useful comments

and corrections. I would also like to thank Rob Calver for several suggestions that greatly improved the book.

Thomas A. Severini
severini@northwestern.edu
Evanston, Illinois

Author

Thomas A. Severini is currently professor of statistics at Northwestern University. His research areas include likelihood inference, nonparametric and semiparametric methods, and applications to econometrics. He is also the author of *Likelihood Methods in Statistics*, *Elements of Distribution Theory*, and *Introduction to Statistical Methods for Financial Models*. He received his PhD in statistics from the University of Chicago. He is a fellow of the American Statistical Association.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1

Introduction

1.1 Analytic Methods

Analytic methods use data to draw conclusions and make decisions. The challenge in using these methods is that the messages of the data are not always clear, and it is often necessary to filter out the noise to see the underlying relationships. Therefore, one distinguishing feature of analytic methods is that they recognize the inherent randomness of data and they are designed to extract useful information in the presence of this randomness.

This is particularly important when analyzing sports data because we all know that the results of a game or other sporting event depend not only on the skill of the participants but also on "luck" and randomness, and separating the contribution of skill from that of luck is not always easy. A second difficulty in analyzing sports data is that a sporting event is a type of "observational study," a study in which important aspects are not under the control of the analyst; that is, we simply observe the data as they are generated, and in contrast to a controlled experiment, we cannot choose which players or teams participate in a particular event or in a given situation.

Given the emphasis on data analysis, it is not surprising that statistical concepts are central to methods presented in this book. Although statistical methodology is a vast topic, fortunately, there are a few central concepts and basic methods that can greatly improve our understanding of data and the processes that generated them. Knowledge in this area will be beneficial to all serious sports fans, whether they simply want to better understand the "new statistics" that have been proposed or whether they want to conduct their own statistical analyses.

There are at least two important roles of statistics in analytic methods. One is the use of statistical methodology designed to efficiently extract relevant information about measurements and their relationships. Statistical models are essential in this process. The models used to analyze sports data are generally empirical rather than based on some underlying theory. That is, the models used describe general features of relationships between variables; such models have been found useful in many fields and form the core of statistical methodology. What distinguishes statistical models from other types of models is the use of the concept of probability in describing relationships; such models are often described as "stochastic," as opposed to a "deterministic" model,

in which probability does not play a role. However, statistical methods do more than recognize randomness; they filter it out, exposing the meaningful relationships in data.

When using statistical models of any type, it is important to keep in mind that they involve some idealization and simplification of complicated physical relationships. However, this does not diminish their usefulness; in fact, it is an essential part of it. Appropriate simplification is a crucial step in stripping away the randomness that often clouds our perception of the salient factors driving sports results.

A second role of statistical concepts in analytic methods is to give a framework for using probability to describe uncertainty. Given the random nature of the results of sporting events, any conclusions we draw from analyzing sports data will naturally have some uncertainty associated with them. Statistical methodology is designed to assess this uncertainty and express it in a useful and meaningful way. Explicit recognition of the random nature of sports is one of the primary contributions of analytic methods, such as those used in sabermetrics, to the analysis of sports data.

1.2 Organization of the Book

Although a central theme of the book is the use of statistical models in understanding and interpreting sports data, before presenting the details of these methods, it is important to understand the basic properties of data. These properties are the subject of Chapter 2, which covers the fundamental methods of describing and summarizing data.

As noted in the previous section, the use of probability theory and statistical methodology to describe relationships and express conclusions is a crucial part of analytic methods. Chapter 3 covers those aspects of probability theory that are necessary to understand the randomness inherent in sports data. These concepts are applied to a number of scenarios in sports in which consideration of the underlying probabilities leads to useful insights. As noted previously, appreciating and understanding randomness is one of the main contributions of analytic methods.

Chapter 4 has several goals. One is to describe the statistical reasoning that underlies the analytic methods described in this book. Another is to present some basic statistical concepts, such as the *margin of error* and *statistical significance*, that play a central role in dealing with the randomness of sports data. Finally, Chapter 4 covers some basic statistical methods that are essential in studying sports data.

Chapters 5 through 7 develop the core statistical procedures for analyzing data based on sports results. Chapter 5 is concerned with detecting the presence of a relationship between variables and measuring the strength of such

relationships. Several different methods are presented, designed to deal with different types of data and different goals for the analysis.

Chapter 6 takes the basic theme of Chapter 5—the relationship between variables—and goes a step further, covering methods for summarizing the relationship between two variables in a concise and useful way. These methods, known collectively as *linear regression,* use statistical methodology to find a function relating the two variables. The simplest method of this type yields a linear function for the variables; Chapter 6 also covers more sophisticated methods that are used when the relationship is nonlinear.

In Chapter 7, these methods are extended to the case of several variables when we wish to describe one of the variables, known as the *response variable*, in terms of the others, known as *predictors*. These methods, also known as linear regression, are perhaps the most commonly used statistical procedures, with applications in a wide range of scientific fields. Chapter 7 contains a detailed discussion of the basic methodology, along with more advanced topics such as the use of categorical variables as predictors, methods for finding the most important predictor, and interaction, which occurs when the effect of one of the predictors depends on the values of other predictors. In addition to the descriptions of the relevant statistical methodology, Chapters 6 and 7 include important information on the strengths and limitations of these methods as well as on the implementation of the methodology and the interpretation of the results.

Chapter 8 discusses some more advanced methods that build on the topics covered in Chapters 5 through 7. Many of these methods are extensions of the regression methodology covered in Chapters 6 and 7, such as logistic regression for modeling the relationship between a binary response variable and predictor variables, and spline models for modeling highly nonlinear relationships. Other methods, such as using pooling to estimate team- and player-specific parameters, principal components analysis for summarizing data, and the use of random effects to analyze variability, introduce new concepts.

The topics covered in this book are similar to those that would be covered in courses on statistical methodology. However, they have been chosen specifically because of their importance and usefulness in analyzing sports data. Therefore, statistical methods that are not useful in analyzing sports data are not covered. Furthermore, many of the topics that are discussed are fairly advanced in the sense that they would not typically be covered in an introductory statistics course.

1.3 Data

The analytic methods described in this book have as their ultimate goal the analysis of data. Therefore, throughout the book, the methodology presented

is illustrated on genuine data drawn from a wide variety of sports. Readers are encouraged to replicate these analyses, as well as to conduct related analyses using their own data.

There is no shortage of data available on the internet. Some sites that have been found useful include sites operated by sports leagues or organizations (e.g., MLB.com, NFL.com, PGA.com, etc.); the sites operated by news organizations (e.g., ESPN.com, SI.com, yahoo.sports.com, etc.); the sports reference sites (e.g., Baseball-Reference.com, Pro-Football-Reference.com); and independent sites such as FanGraphs.com. The "Reference.com" sites (e.g., Baseball-Reference.com and Pro-Football-Reference.com) are particularly noteworthy for the detailed data available and for their search engines, which are invaluable for finding data relevant to a specific question. Although some of these features require a modest subscription fee, for serious study of sports, the cost is minor.

For many of the examples given in this book, there are several possible sources for the data, and a specific source is not given. This is generally the case when the data analyzed are based directly on game results. Although all sites for a given sport contain the same basic data, different sites often contain different specialized data, such as "splits," data for specific situations, or "advanced statistics" that have been calculated from the basic data. Also, it is sometimes easier to retrieve or download data from some sites than others, depending on the user's software and experience. Therefore, it might be helpful to check a number of sources to find the ones that are most useful. When the data under consideration are more site specific, a reference to the source of the data is given.

Many of the datasets analyzed in this book are available in Excel format by following the link on my website (`http://www.taseverini.com`). When a dataset is available, it is denoted in the text by a label of the form Dataset C.x, where C denotes the relevant chapter and x denotes the dataset within that chapter; for example, the first available dataset in Chapter 2 is denoted as Dataset 2.1. A list of all available datasets, together with a description of the variables in the Excel spreadsheet and the source of the data, is given at the end of the book.

1.4 Computation

Nearly all the methods presented in this book require computation, and for the vast majority, some type of software is needed. Two options are discussed here. One is Excel and Chapters 2 through 7 each contain a section outlining how to perform the calculations in Excel, enabling readers to repeat the analyses given in the book. These sections assume that the reader has basic knowledge of standard Excel functions; for others, there are many excellent books on

Excel available. The commands listed in these sections are based on the 2010 version of Excel.

Statistical calculations in Excel require the Analysis ToolPak, which is available but is not automatically loaded when Excel is started. To load the Analysis ToolPak, use the following procedure, which is taken from the Excel help file.

1. Click the **File** tab, click **Options**, and then click the **Add-Ins** category.
2. In the **Manage** box, select **Excel Add-Ins** and then click **Go**.
3. In the **Add-Ins available** box, select the **Analysis ToolPak** check box and then click **OK**.

Tip: If **Analysis ToolPak** is not listed in the **Add-Ins available** box, click **Browse** to locate it. If you are prompted that the Analysis ToolPak is not currently installed on your computer, click **Yes** to install it. The **Data Analysis** button will now be available under the **Data** tab.

A second option is R, a comprehensive computing environment that can be used for a wide range of numerical calculations, including all of the statistical analyses described in this book. It can also be used to produce graphs and it can function as a programming language. One author describes R as an "overgrown calculator" (Daalgaard, 2008) and, for the way R is used in this book, that is a good way to think about it. R is available, free of charge, for a wide range of platforms, including Microsoft Windows and Apple macOS, at https://www.r-project.org. Many R users find it convenient to run R using RStudio, an open source Integrated Development Environment that has a number of useful features. RStudio is available at https://www.rstudio.com, free of charge. Implementation of the methodology discussed in this book using R is discussed in each of Chapters 2 through 8.

Although the use of R in this book is aimed at those without much R experience, it does not contain the type of detailed information on the basics of R which beginners will find useful. Hence, such readers will benefit from a more extensive introduction. Fortunately, there are several useful references available on the internet, including https://www.statmethods.net/r-tutorial/index.html, which contains an R tutorial; a useful beginner's guide is available on https://www.computerworld.com, easily found by googling "computerworld introduction to R." There are also many good books available on using R for statistical analysis, including *Introductory Statistics with R* by Peter Dalgaard (Dalgaard, 2008) and *R for Data Science* by Garrett Grolemund and Hadley Wickham, which is available at https://r4ds.had.co.nz/index.html.

1.5 Suggestions for Further Reading

Each chapter concludes with a section containing references and suggestions for further reading, as well as some occasional comments on the material covered in the main text.

Analysis of sports data is a rapidly growing field, and there are many books describing specific analyses. Here I give a few examples; there are many others. The work of Tango, Lichtman, and Dolphin (2007) contains a detailed treatment of how analytic methods can be used to answer specific questions about baseball strategy. Keri's (2006) work is a collection of interesting essays in which analytic methods are used to study issues ranging from a comparison of Babe Ruth to Barry Bonds (Silver, 2006a) to what baseball statistics can tell us about steroids (Silver, 2006c). Joyner (2008) provides similar analyses for football-related questions. The works of Moskowitz and Wertheim (2011) and Winston (2009) both show how analytic methods can be used to address a wide range of sports issues.

The mathematical and statistical topics discussed in this book are covered in many books on math and statistics; specific references are given in the relevant chapters. The work of Cox and Donnelly (2011, Chapter 6) contains a detailed discussion of the role of statistical models, expanding on the discussion in Section 1.1. Freedman (2009) discusses the theory and application of statistical models, paying particular attention to the analysis of observational data; much of that discussion is relevant to the analysis of sports data. Silver (2012) gives an entertaining, nontechnical introduction to analytic methods from the point of view of prediction; many of the points raised by Silver are useful in the analysis of sports data.

Bibliography

[1] Achen, C. H. (1982). *Interpreting and Using Regression*. Beverly Hills, CA: Sage.

[2] Agresti, A., and Finlay, B. (2009). *Statistical Methods for the Social Sciences*, 4th ed. Upper Saddle River, NJ: Pearson.

[3] Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

[4] Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*, 6th ed. Boca Raton, FL: CRC Press.

[5] Click, J. (2006). What if Rickey Henderson had Pete Incaviglia's legs? In J. Keri (Ed.), *Baseball between the Numbers: Why Everything You Know about the Game Is Wrong* (pp. 112–126). New York: Basic Books.

[6] Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24**, 295–313.

[7] Cox, D. R., and Donnelly, C. A. (2011). *Principles of Applied Statistics*. New York: Cambridge University Press.

[8] Cox, D. R., and Solomon, P. J. (2003). *Components of Variance*. Boca Raton, FL: CRC Press.

[9] Dalgaard, P. (2008). *Introductory Statistics with R*, 2nd ed. New York: Springer.

[10] de Oliveira, S. P. (2017). A very basic introduction to Random Forests using R. Retrieved from https://www.blopig.com/blog/2017/04/a-very-basic-introduction-to-random-forests-using-r/.

[11] Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford, UK: Oxford University Press.

[12] Dobson, A. J., and Barnett, A. G. (2008). *An Introduction to Generalized Linear Models*, 3rd ed. Boca Raton, FL: CRC Press.

[13] Everitt, B., and Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. New York: Springer.

[14] Freedman, D. A. (1999). Ecological inference and the ecological fallacy. Retrieved from http://www.stanford.edu/class/ed260/freedman549.pdf.

[15] Freedman, D. A. (2009). *Statistical Models: Theory and Practice*. New York: Cambridge University Press.

[16] Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

[17] Gilovich, T. (1991). *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*. New York: Free Press.

[18] Goldner, K. (2012). A Markov model of football: using stochastic processes to model a football drive. *Journal of Quantitative Analysis in Sports* **8**, 1–18.

[19] Goldstein, D. (2013). Stadium/home team effects in making field goals. Retrieved from http://www.decisionsciencenews.com/2013/02/13/stadium-home-team-effects-in-making-field-goals/.

[20] Goldstein, H. (2011). *Multilevel Statistical Models*, 4th ed. Chichester, UK: Wiley.

[21] Grinstead, C. M., and Snell, J. L. (1997). *Introduction to Probability*, 2nd rev. ed. Providence, RI: American Mathematical Society.

[22] Huff, D. (1993). *How to Lie with Statistics*. New York: Norton.

[23] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

[24] Joyner, K. C. (2008). *Blindsided: Why the Left Tackle Is Overrated and Other Contrarian Football Thoughts*. Hoboken, NJ: Wiley.

[25] Keri, J. (Ed.) (2006). *Baseball between the Numbers: Why Everything You Know About the Game Is Wrong*. New York: Basic Books.

[26] Kuhn, M., and Johnson, K. (2016). *Applied Predictive Modeling*. New York: Springer.

[27] Lederer, R. (2009). Using *Z*-scores to rank pitchers. Retrieved from http://baseballanalysts.com/archives/2009/02/using_zscores_t.php.

[28] Lependorf, D. (2012). Is there a better way of comparing players between historical eras? Retrieved from http://www.hardballtimes.com/main/article/is-there-a-better-way-of-comparing-players-between-historical-eras/.

[29] McClave, J. T., and Sincich, T. (2006). *A First Course in Statistics*, 9th ed. Upper Saddle River, NJ: Pearson.

[30] Miller, A. J. (2002). *Subset Selection in Regression*, 2nd ed. Boca Raton, FL: CRC Press.

[31] Mlodinow, L. (2008). *The Drunkard's Walk: How Randomness Rules Our Lives*. New York:Vintage Books.

[32] Moore, D., and McCabe, G. (2005). *Introduction to the Practice of Statistics*, 4th ed. New York: Freeman.

[33] Moskowitz, T. J., and Wertheim, L. J. (2011). *Scorecasting: The Hidden Influences Behind How Sports Are Played and Games Are Won*. New York: Three Rivers Press.

[34] Mosteller, F., and Tukey, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics*. Upper Saddle River, NJ: Pearson.

[35] Pinheiro, J., and Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

[36] Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In T. Pukkila and S. Puntanen (Eds.), *Proceedings of the Second International Tampere Conference in Statistics*, pp. 245–250. Department of Mathematical Sciences/Statistics, University of Tampere, Tampere, Finland.

[37] Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* **15**, 351–357.

[38] Rodgers, J. L., and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician* **42**, 59–66.

[39] Ross, S. (2006). *A First Course in Probability*, 7th ed. Upper Saddle River, NJ: Pearson.

[40] Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

[41] Schilling, M. F. (1990). The longest run of heads. *The College Mathematics Journal* **21**, 196–207.

[42] Silver, N. (2006a). Batting practice: Is Barry Bonds better than Babe Ruth? In J. Keri (Ed.), *Baseball between the Numbers: Why Everything You Know about the Game Is Wrong* (pp. xxxvii–lxii). New York: Basic Books.

[43] Silver, N. (2006b). Is David Ortiz a clutch hitter? In J. Keri (Ed.), *Baseball between the Numbers:Why Everything You Know about the Game Is Wrong* (pp. 14–34). New York: Basic Books.

[44] Silver, N. (2006c). What do statistics tell us about steroids? In J. Keri (Ed.), *Baseball between the Numbers: Why Everything You Know about the Game Is Wrong* (pp. 326–342). New York: Basic Books.

[45] Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't*. New York: Penguin Press.

[46] Snedecor, G. W., and Cochran, W. G. (1980). *Statistical Methods*, 7th ed. Ames: Iowa State University Press.

[47] Tango, T. M., Lichtman, M. G., and Dolphin, A. E. (2007). *The Book: Playing the Percentages in Baseball*. Washington, DC: Potomac Books.

[48] Thomas, D. R., Hughes, E., and Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research* **45**, 253–275.

[49] Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Cheshire, UK: Graphics Press.

[50] Verducci, T. (2013). Virtue, and victory, no longer synonymous with patience at the plate. Retrieved from http://sportsillustrated.cnn.com/mlb/news/20130423/joey-vottojayson-werth-taking-pitches/.

[51] Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics* **14**, 121–140.

[52] Wardell, P. (2011). 20 statistical oddities from the 2011 MLB season so far. Retrieved from http://bleacherreport.com/articles/687498-20-statistical-oddities-from-the-2011-season-so-far.

[53] Winston, W. L. (2009). *Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football*. Princeton, NJ: Princeton University Press.

[54] Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*, 2nd ed. Boca Raton, FL: CRC Press.

[55] Woolner, K. (2006). Are teams letting their closers go to waste? In J. Keri (Ed.), *Baseball between the Numbers: Why Everything You Know about the Game Is Wrong* (pp. 58–73). New York: Basic Books.