

Preface to the First Edition

Over twenty years, this author has been in the enterprise of teaching introductory statistics to an audience that is taking the class to satisfy their mathematics requirement. This is a challenging endeavor because the students have little prior knowledge about the discipline of statistics and many of them are anxious about mathematics and computation. Statistical concepts and examples are usually presented in a particular context. However, one obstacle in teaching this introductory class is that we often describe the statistical concepts in a context, such as medicine, law, or agriculture that is completely foreign to the undergraduate student. The student has a much better chance of understanding concepts in probability and statistics if they are described in a familiar context.

Many students are familiar with sports either as a participant or a spectator. They know of the popular athletes, such as Tiger Woods and Barry Bonds, and they are generally knowledgeable with the rules of the major sports, such as baseball, football, and basketball. For many students sports is a familiar context in which an instructor can describe statistical thinking.

The goal of this book is to provide a collection of examples and exercises applying probability and statistics to the sport of baseball. Why baseball instead of other sports?

- Baseball is the great American game. Baseball is great in that it has a rich history of teams and players, and many people are familiar with the basic rules of the game. The popularity of baseball is reflected by the large number of movies that have been produced about baseball teams and players.
- Baseball is the most statistical of all sports. Hitters and pitchers are identified by their corresponding hitting and pitching statistics. For example, Babe Ruth is forever identified by the statistic 60, which was the number of home runs hit in his 1927 season. Bob Gibson is famous for his record low earned run average of 1.12 during the 1968 season. A flood of different statistical measures are used to rate players and salaries of players are determined in part by these statistics. There is an active effort among baseball writers to learn more about baseball issues by using statistics.
- A wealth of baseball data is currently available over the Internet. Player and team hitting and pitching statistics can be easily found. Comparisons between players of different eras can be made using a downloadable dataset that gives hitting and pitching data for all players who have ever played professional baseball.

This book is organized using the same basic organization structure presented in most introductory statistics texts. After an introductory chapter, there is a chapter on the analysis on a single batch of data, followed by chapters on the comparison of batches, and the analysis of relationships. There are chapters on introductory and more advanced topics in probability, followed by topics in statistical inference. Each chapter contains a number of essays or case studies that describe the analysis of statistical or probabilistic methods to particular baseball data sets. After the collection of case studies in each chapter, there is a set of activities and exercises that suggest further exploration of baseball datasets similar to the analysis presented in the case studies.

How can this book be used in teaching probability or statistics? We suggest several uses of this material.

- This book can be used as the framework for a one-semester introductory statistics class that is focused on baseball. Such a class has been taught at the author's home institution. This course covers the basic topics of a beginning statistics course (data analysis, introductory probability, and concepts of inference) using baseball as the primary source of applications. This course is suitable for students who are interested or curious about the game of baseball. It is also suitable for students with sports-related majors, such as sports management or sports medicine.
- This book can also be used as a resource for instructors who wish to infuse their present course in probability or statistics with applications from baseball. The material in this book has been presented at different levels to make it useable for introductory and more advanced courses. The case studies can be used by the instructor to present the particular topic within a baseball context and then the associated exercises and activities can be used for homework. The case studies can serve as useful springboards for undergraduate students who wish to do additional explorations on baseball data.

Acknowledgements

I am appreciative of the support given to this project by the Division of Undergraduate Education of the National Science Foundation and by my colleagues in the Department of Mathematics and Statistics at Bowling Green State University. The text was used for a number of experimental sections of MATH 115 Introduction to Statistics at Bowling Green State University and I am grateful for the valuable feedback from the students who enrolled in this course. In addition, Chris Andrews, Jay Bennett, Eric Bradlow, Jim Cochran, Joe Gallian, Carl Morris, Jerome Reiter, Ken Ross, Steve Samuels, Bob Wardrop, and Dex Whittinghill provided many helpful suggestions in reviewing the book. I thank the editors of the MAA publications for their support, particularly Zaven Karien and Dave Kullman. Last, but certainly not least, I thank my wife Anne, and children Lynne, Bethany, and Steven for their understanding and great patience during the completion of this text.

Jim Albert
December 2002

Preface to the Second Edition

The author has been gratified with the reception to the first edition of this text. It has been enjoyable teaching introductory statistics from a baseball viewpoint, as one can discuss statistical concepts in the context of current and historical players and teams.

Sports provide a wonderful context for teaching introductory statistics and other texts have been recently published using a sports theme. Rothman (2012) is a comprehensive statistics text with a baseball theme, and Tabor and Franklin (2011) is a statistics text with examples from a wide range of sports. There is a wealth of publicly available baseball data and Marchi and Albert (2013) describe the use of these datasets together with the open-source statistics software system R (R Core Team (2015)) to implement a variety of baseball studies.

Since the date of the first edition, there have been changes both in the players who play baseball, and also in the development and use of analytics. So this motivates revisions to the text reflecting these changes.

In this edition, many of the case studies and exercises have been revised to use data from current teams and players. I encourage the instructor to always use current season data since the modern players and teams are most familiar to students.

Also exercises have been added to the chapters reflecting some of the newer types of baseball data. The pitchFX system has been tracking the trajectories of baseball pitches since 2006 and we know more about the breaks, locations, and speeds of pitches. Websites such as fangraphs.com contain much of this pitchFX data for both batters and pitchers and this website also gives information about the location, speed, and type (flyball, pop-up, and grounders) of balls put in play. The tracking technology Statcast contains information about the speed of runners and fielders and velocities of balls coming off a bat.

To facilitate the use of the data described in the examples and exercises, all of the datasets are currently available using the StatCrunch statistical software system published by Pearson (in StatCrunch search using the acronym TSUB and the example or exercise number). In addition, the datasets are available as text files at <https://bayesball.github.io/>.

Jim Albert
December 2016