CONTENTS

Preface xi
Acknowledgments xv
Abbreviations xvii

PART I BASEBALL

1. Baseball’s Pythagorean Theorem 3
2. Who Had a Better Year: Mike Trout or Kris Bryant? 12
3. Evaluating Hitters by Linear Weights 18
4. Evaluating Hitters by Monte Carlo Simulation 31
5. Evaluating Baseball Pitchers, Forecasting Future Pitcher Performance, and an Introduction to Statcast 44
6. Baseball Decision Making 60
7. Evaluating Fielders 73
8. Win Probability Added (WPA) 84
9. Wins Above Replacement (WAR) and Player Salaries 92
10. Park Factors 101
11. Streakiness in Sports 105
12. The Platoon Effect 124
13. Was Tony Perez a Great Clutch Hitter? 127
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.</td>
<td>Pitch Count, Pitcher Effectiveness, and PITCHf/x Data</td>
<td>133</td>
</tr>
<tr>
<td>15.</td>
<td>Would Ted Williams Hit .406 today?</td>
<td>139</td>
</tr>
<tr>
<td>16.</td>
<td>Was Joe DiMaggio’s 56-Game Hitting Streak the Greatest Sports Record of All Time?</td>
<td>142</td>
</tr>
<tr>
<td>17.</td>
<td>Projecting Major League Performance</td>
<td>151</td>
</tr>
</tbody>
</table>

**PART II FOOTBALL**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.</td>
<td>What Makes NFL Teams Win?</td>
<td>159</td>
</tr>
<tr>
<td>19.</td>
<td>Who’s Better: Brady or Rodgers?</td>
<td>164</td>
</tr>
<tr>
<td>20.</td>
<td>Football States and Values</td>
<td>170</td>
</tr>
<tr>
<td>21.</td>
<td>Football Decision Making 101</td>
<td>178</td>
</tr>
<tr>
<td>22.</td>
<td>If Passing Is Better than Running, Why Don’t Teams Always Pass?</td>
<td>186</td>
</tr>
<tr>
<td>23.</td>
<td>Should We Go for a One-Point or a Two-Point Conversion?</td>
<td>195</td>
</tr>
<tr>
<td>24.</td>
<td>To Give Up the Ball Is Better than to Receive: The Case of College Football Overtime</td>
<td>207</td>
</tr>
<tr>
<td>25.</td>
<td>Has the NFL Finally Gotten the OT Rules Right?</td>
<td>211</td>
</tr>
<tr>
<td>27.</td>
<td>Player Tracking Data in the NFL</td>
<td>229</td>
</tr>
</tbody>
</table>

**PART III BASKETBALL**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.</td>
<td>Basketball Statistics 101: The Four Factor Model</td>
<td>249</td>
</tr>
<tr>
<td>29.</td>
<td>Linear Weights for Evaluating NBA Players</td>
<td>259</td>
</tr>
<tr>
<td>30.</td>
<td>Adjusted +/- Player Ratings</td>
<td>265</td>
</tr>
<tr>
<td>31.</td>
<td>ESPN RPM and FiveThirtyEight RAPTOR Ratings</td>
<td>282</td>
</tr>
</tbody>
</table>
### Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>NBA Lineup Analysis</td>
<td>289</td>
</tr>
<tr>
<td>33</td>
<td>Analyzing Team and Individual Matchups</td>
<td>296</td>
</tr>
<tr>
<td>34</td>
<td>NBA Salaries and the Value of a Draft Pick</td>
<td>303</td>
</tr>
<tr>
<td>35</td>
<td>Are NBA Officials Prejudiced?</td>
<td>307</td>
</tr>
<tr>
<td>36</td>
<td>Pick-n-Rolling to Win, the Death of Post Ups and Isos</td>
<td>313</td>
</tr>
<tr>
<td>37</td>
<td>SportVU, Second Spectrum, and the Spatial Basketball Data Revolution</td>
<td>321</td>
</tr>
<tr>
<td>38</td>
<td>In-Game Basketball Decision Making</td>
<td>341</td>
</tr>
<tr>
<td></td>
<td><strong>PART IV OTHER SPORTS</strong></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Soccer Analytics</td>
<td>355</td>
</tr>
<tr>
<td>40</td>
<td>Hockey Analytics</td>
<td>373</td>
</tr>
<tr>
<td>41</td>
<td>Volleyball Analytics</td>
<td>385</td>
</tr>
<tr>
<td>42</td>
<td>Golf Analytics</td>
<td>391</td>
</tr>
<tr>
<td>43</td>
<td>Analytics and Cyber Athletes: The Era of e-Sports</td>
<td>398</td>
</tr>
<tr>
<td></td>
<td><strong>PART V SPORTS GAMBLING</strong></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Sports Gambling 101</td>
<td>409</td>
</tr>
<tr>
<td>45</td>
<td>Freakonomics Meets the Bookmaker</td>
<td>420</td>
</tr>
<tr>
<td>46</td>
<td>Rating Sports Teams</td>
<td>423</td>
</tr>
<tr>
<td>47</td>
<td>From Point Ratings to Probabilities</td>
<td>447</td>
</tr>
<tr>
<td>48</td>
<td>The NCAA Evaluation Tool (NET)</td>
<td>464</td>
</tr>
<tr>
<td>49</td>
<td>Optimal Money Management:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Kelley Growth Criterion</td>
<td>468</td>
</tr>
<tr>
<td>50</td>
<td>Calcuttas</td>
<td>474</td>
</tr>
</tbody>
</table>
PART VI METHODS AND MISCELLANEOUS

51. How to Work with Data Sources: Collecting and Visualizing Data 479
52. Assessing Players with Limited Data: The Bayesian Approach 490
53. Finding Latent Patterns through Matrix Factorization 499
54. Network Analysis in Sports 508
55. Elo Ratings 524
56. Comparing Players from Different Eras 531
57. Does Fatigue Make Cowards of Us All? The Case of NBA Back-to-Back Games and NFL Bye Weeks 538
58. The College Football Playoff 543
59. Quantifying Sports Collapses 551
60. Daily Fantasy Sports 559

Bibliography 569
Index 579
CHAPTER 1

BASEBALL’S PYTHAGOREAN THEOREM

The more runs that a baseball team scores, the more games the team should win. Conversely, the fewer runs a team gives up, the more games the team should win. Bill James, probably the most celebrated advocate of applying mathematics to analysis of Major League Baseball (often called sabermetrics), studied many years of Major League Baseball standings and found that the percentage of games won by a baseball team can be well approximated by the formula

\[
\frac{\text{runs scored}^2}{\text{runs scored}^2 + \text{runs allowed}^2} = \text{Estimate of percentage of games won.}
\]  

(1)

This formula has several desirable properties:

- Predicted win percentage is always between 0 and 1.
- An increase in runs scored increases predicted win percentage.
- A decrease in runs allowed increases predicted win percentage.

Consider a right triangle with a hypotenuse (the longest side) of length c and two other sides of length a and b. Recall from high school geometry that the Pythagorean Theorem states that a triangle is a right triangle if and only if \(a^2 + b^2 = c^2\) must hold. For example, a
triangle with sides of lengths 3, 4, and 5 is a right triangle because $3^2 + 4^2 = 5^2$. The fact that equation (1) adds up the squares of two numbers led Bill James to call the relationship described in (1) Baseball’s Pythagorean Theorem.

Let’s define $\frac{\text{runs scored}}{\text{runs allowed}}$ as a team’s scoring ratio. If we divide the numerator and denominator of (1) by $(\text{runs allowed})^2$, then the value of the fraction remains unchanged and we may re-write (1) as equation (1’).

$$\frac{R^2}{R^2 + 1} = \text{Estimate of percentage of games won}$$

(1’)

Figure 1-1 (see file Mathleticschapter1files.xlsx for all of this chapter’s analysis) shows how well (1’) predicts teams’ winning percentages for Major League Baseball teams during the 2005–2016 seasons. For example, the 2016 Los Angeles Dodgers scored 725 runs and gave up 638 runs. Their scoring ratio was $\frac{725}{638} = 1.136$. Their predicted win percentage from Baseball’s Pythagorean Theorem was $\frac{1.136^2}{1.136^2 + 1} = .5636$. The 2016 Dodgers actually won a fraction $\frac{91}{162} = .5618$ of their games. Thus (1’) was off by 0.18% in predicting the percentage of games won by the Dodgers in 2016.

For each team define Error in Win Percentage Prediction to equal Actual Winning Percentage minus Predicted Winning Percentage. For example, for the 2016 Atlanta Braves, Error $=.42 - .41 = .01$ (or 1.0%), and for the 2016 Colorado Rockies, Error $=.46 - .49 = -.03$ (or 3%). A positive error means that the team won more games than predicted while a negative error means the team won fewer games than predicted. Column J computes for each team the absolute value of the prediction error. Recall that absolute value of a number is simply the distance of the number from 0. That is, $|5| = |-5| = 5$. In cell J1 we average the absolute prediction errors for each team to obtain a measure of how well our predicted win percentages fit the actual team winning percentages. The average of absolute forecasting
Baseball’s Pythagorean Theorem

5

errors is called the MAD (mean absolute deviation). We find that for our dataset the predicted winning percentages of the Pythagorean Theorem were off by an average of 2.17% per team.

Instead of blindly assuming win percentage can be approximated by using the square of the scoring ratio, perhaps we should try a formula to predict winning percentage, such as

\[ \frac{R^{\exp}}{R^{\exp} + 1}. \]

If we vary \( \exp \) in (2) we can make (2) better fit the actual dependence of winning percentage on the scoring ratio for different sports.

1. Why didn’t we just average the actual errors? Because averaging positive and negative errors would result in positive and negative errors canceling out. For example, if one team wins 5% more games than (1’’) predicts and another team wins 5% less games than (1’’) predicts, the average of the errors is 0 but the average of the absolute errors is 5%. Of course, in this simple situation estimating the average error as 5% is correct while estimating the average error as 0% is nonsensical.
For baseball, we will allow exp in (2) (exp is short for exponent) to vary between 1 and 3. Of course exp = 2 reduces to the Pythagorean Theorem.

Figure 1-2 shows how the MAD changes as we vary exp between 1 and 3. This was done using the Data Table feature in Excel. We see that indeed exp = 1.8 yields the smallest MAD (1.99%). An exp value of 2 is almost as good (MAD of 2.05%), so for simplicity we will stick with Bill James’s view that exp = 2. Therefore exp = 2 (or 1.8) yields the best forecasts if we use an equation of form (2). Of course, there might be another equation that predicts winning percentage better than the Pythagorean Theorem from runs scored and allowed. The Pythagorean

2. See Chapter 1 Appendix for an explanation of how we used Data Tables to determine how MAD changes as we vary exp between 1 and 3. Additional information available at https://support.office.com/en-us/article/calculate-multiple-results-by-using-a-data-table-e95e2487-6ca6-4413-ad12-77542a5ea50b.
Baseball’s Pythagorean Theorem

rean Theorem is simple and intuitive, however, and does very well. After all, we are off in predicting team wins by an average of 162 * .0205, which is approximately three wins per team. Therefore, I see no reason to look for a more complicated (albeit slightly more accurate) model.

**HOW WELL DOES THE PYTHAGOREAN THEOREM FORECAST?**

To test the utility of the Pythagorean Theorem (or any prediction model) we should check how well it forecasts the future. We chose to compare the Pythagorean Theorem’s forecast for each Major League Baseball playoff series (2005–2016) against a prediction based just on games won. For each playoff series the Pythagorean method would predict the winner to be the team with the higher scoring ratio while the “games won” approach simply predicts the winner of a playoff series to be the team that won more games. We found that the Pythagorean approach correctly predicted 46 of 84 playoff series (54.8%) while the “games won” approach correctly predicted the winner of only 55% (44 out of 80) playoff series. The reader is probably disappointed that even the Pythagorean method only correctly forecasts the outcome of under 54% of baseball playoff series. We believe that the regular season is a relatively poor predictor of the playoffs in baseball because a team’s regular season record depends a lot on the performance of five starting pitchers. During the playoffs, teams only use three or four starting pitchers, so a lot of the regular season data (games involving the fourth and fifth starting pitchers) are not relevant for predicting the outcome of the playoffs.

For anecdotal evidence of how the Pythagorean Theorem forecasts the future performance of a team better than a team’s win-loss record, consider the case of the 2005 Washington Nationals. On July 4, 2005, the Nationals were in first place with a record of 50–32. If we had extrapolated this win percentage, we would have predicted

3. In four playoff series the opposing teams had identical win-loss records, so the “games won” approach could not make a prediction.
a final record of 99–63. On July 4, 2005, the Nationals’ scoring ratio was .991. On July 4, 2005, equation (1) would predict the Nationals to win around half (40) of the remaining 80 games and finish with a 90–72 record. In reality, the Nationals only won 31 of their remaining games and finished at 81–81!

**IMPORTANCE OF PYTHAGOREAN THEOREM**

The Baseball Pythagorean Theorem is also important because it allows us to determine how many extra wins (or losses) will result from a trade. As an example, suppose a team has scored 850 runs during a season and also given up 800 runs. Suppose we trade an SS (Joe) who “created” 150 runs for a shortstop (Greg) who created 170 runs in the same number of plate appearances. This trade will cause the team (all other things being equal) to score 170−150 = 20 more runs. Before the trade, $R = \frac{850}{800} = 1.0625$, and we would predict the team to have won \( \frac{162 \times 1.0625^2}{1+1.0625^2} = 85.9 \) games.

After the trade, $R = \frac{870}{800} = 1.0875$, and we would predict the team to have won \( \frac{162 \times 1.0875^2}{1+1.0875^2} = 87.8 \) games. Therefore, we estimate the trade makes our team 87.8−85.9 = 1.9 games better. In Chapter 9, we will see how the Pythagorean Theorem can be used to help determine fair salaries for Major League Baseball players.

**FOOTBALL AND BASKETBALL**

"PYTHAGOREAN THEOREMS"

Does the Pythagorean Theorem hold for football and basketball? Daryl Morey, currently the General Manager for the Houston Rockets NBA team, has shown that for the NFL, equation (2) with

4. In Chapters 2–4 we will explain in detail how to determine how many runs a hitter creates.
exp = 2.37 gives the most accurate predictions for winning percentage, while for the NBA, equation (2) with exp = 13.91 gives the most accurate predictions for winning percentage. Figure 1-3 gives the predicted and actual winning percentages for the 2015 NFL, while Figure 1-4 gives the predicted and actual winning percentages for the 2015–2016 NBA. See the file Sportshw1.xls.

For the 2008–2015 NFL seasons we found MAD was minimized by exp = 2.8. Exp = 2.8 yielded a MAD of 6.08%, while Morey's exp = 2.37 yielded a MAD of 6.39%. For the NBA seasons 2008–2016 we found exp = 14.4 best fit actual winning percentages. The MAD for these seasons was 2.84% for exp = 14.4 and 2.87% for exp = 13.91. Since Morey’s values of exp are very close in accuracy to the values we found from recent seasons we will stick with Morey’s values of exp. See file Sportshw1.xls.

Assuming the errors in our forecasts follow a normal random variable (which turns out to be a reasonable assumption) we would...
Chapter 1

expect around 95% of our NBA win forecasts to be accurate within $2.5 \times \text{MAD} = 7.3\%$. Over 82 games this is about 6 games. So whenever the Pythagorean forecast for wins is off by more than six games, the Pythagorean prediction is an “outlier.” When we spot outliers we try and explain why they occurred. The 2006–2007 Boston Celtics had a scoring ratio of .966, and Pythagoras predicts the Celtics should have won 31 games. They won seven fewer games (24). During that season many people suggested the Celtics “tanked” games to improve their chance of having the #1 pick (Greg Oden and Kevin Durant went 1–2) in the draft lottery. The shortfall in the Celtics’ wins does not prove this conjecture, but the evidence is consistent with the Celtics winning substantially fewer games than chance would indicate.

![FIGURE 1.4](image-url)  
**FIGURE 1.4** Predicted NBA Winning Percentages: $\text{Exp}=13.91$. 

<table>
<thead>
<tr>
<th>Year</th>
<th>Team</th>
<th>Wins</th>
<th>Losses</th>
<th>Points</th>
<th>Opp Points</th>
<th>Ratio</th>
<th>Pred W-L%</th>
<th>Act W-L%</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-16</td>
<td>Atlanta Hawks</td>
<td>48</td>
<td>34</td>
<td>8433</td>
<td>8137</td>
<td>1.04</td>
<td>0.622</td>
<td>0.585</td>
<td>0.037</td>
</tr>
<tr>
<td>2015-16</td>
<td>Boston Celtics</td>
<td>48</td>
<td>34</td>
<td>8669</td>
<td>8406</td>
<td>1.03</td>
<td>0.606</td>
<td>0.585</td>
<td>0.021</td>
</tr>
<tr>
<td>2015-16</td>
<td>Brooklyn Nets</td>
<td>21</td>
<td>61</td>
<td>8089</td>
<td>8692</td>
<td>0.93</td>
<td>0.269</td>
<td>0.256</td>
<td>0.013</td>
</tr>
<tr>
<td>2015-16</td>
<td>Charlotte Hornets</td>
<td>48</td>
<td>34</td>
<td>8476</td>
<td>8256</td>
<td>1.03</td>
<td>0.592</td>
<td>0.585</td>
<td>0.007</td>
</tr>
<tr>
<td>2015-16</td>
<td>Chicago Bulls</td>
<td>42</td>
<td>40</td>
<td>8335</td>
<td>8456</td>
<td>0.99</td>
<td>0.450</td>
<td>0.512</td>
<td>0.062</td>
</tr>
<tr>
<td>2015-16</td>
<td>Cleveland Cavaliers</td>
<td>57</td>
<td>25</td>
<td>8555</td>
<td>8063</td>
<td>1.06</td>
<td>0.695</td>
<td>0.695</td>
<td>6E-05</td>
</tr>
<tr>
<td>2015-16</td>
<td>Dallas Mavericks</td>
<td>42</td>
<td>40</td>
<td>8388</td>
<td>8413</td>
<td>1</td>
<td>0.490</td>
<td>0.512</td>
<td>0.022</td>
</tr>
<tr>
<td>2015-16</td>
<td>Denver Nuggets</td>
<td>33</td>
<td>49</td>
<td>8355</td>
<td>8609</td>
<td>0.97</td>
<td>0.397</td>
<td>0.402</td>
<td>0.005</td>
</tr>
<tr>
<td>2015-16</td>
<td>Detroit Pistons</td>
<td>44</td>
<td>38</td>
<td>8361</td>
<td>8311</td>
<td>1.01</td>
<td>0.521</td>
<td>0.537</td>
<td>0.016</td>
</tr>
<tr>
<td>2015-16</td>
<td>Golden State Warriors</td>
<td>73</td>
<td>9</td>
<td>9421</td>
<td>8539</td>
<td>1.1</td>
<td>0.797</td>
<td>0.89</td>
<td>0.093</td>
</tr>
<tr>
<td>2015-16</td>
<td>Houston Rockets</td>
<td>41</td>
<td>41</td>
<td>8737</td>
<td>8721</td>
<td>1</td>
<td>0.506</td>
<td>0.5</td>
<td>0.006</td>
</tr>
<tr>
<td>2015-16</td>
<td>Indiana Pacers</td>
<td>45</td>
<td>37</td>
<td>8377</td>
<td>8237</td>
<td>1.02</td>
<td>0.558</td>
<td>0.543</td>
<td>0.009</td>
</tr>
<tr>
<td>2015-16</td>
<td>Los Angeles Clippers</td>
<td>53</td>
<td>29</td>
<td>8569</td>
<td>8218</td>
<td>1.04</td>
<td>0.641</td>
<td>0.646</td>
<td>0.005</td>
</tr>
<tr>
<td>2015-16</td>
<td>Los Angeles Lakers</td>
<td>17</td>
<td>65</td>
<td>7982</td>
<td>8766</td>
<td>0.91</td>
<td>0.214</td>
<td>0.207</td>
<td>0.007</td>
</tr>
<tr>
<td>2015-16</td>
<td>Memphis Grizzlies</td>
<td>42</td>
<td>40</td>
<td>8126</td>
<td>8310</td>
<td>0.98</td>
<td>0.423</td>
<td>0.512</td>
<td>0.089</td>
</tr>
<tr>
<td>2015-16</td>
<td>Miami Heat</td>
<td>48</td>
<td>34</td>
<td>8204</td>
<td>8069</td>
<td>1.02</td>
<td>0.557</td>
<td>0.585</td>
<td>0.028</td>
</tr>
<tr>
<td>2015-16</td>
<td>Milwaukee Bucks</td>
<td>33</td>
<td>49</td>
<td>8122</td>
<td>8465</td>
<td>0.96</td>
<td>0.360</td>
<td>0.402</td>
<td>0.042</td>
</tr>
<tr>
<td>2015-16</td>
<td>Minnesota Timberwolves</td>
<td>29</td>
<td>53</td>
<td>8398</td>
<td>8688</td>
<td>0.97</td>
<td>0.384</td>
<td>0.354</td>
<td>0.03</td>
</tr>
<tr>
<td>2015-16</td>
<td>New Orleans Pelicans</td>
<td>30</td>
<td>52</td>
<td>8423</td>
<td>8734</td>
<td>0.96</td>
<td>0.377</td>
<td>0.366</td>
<td>0.011</td>
</tr>
<tr>
<td>2015-16</td>
<td>New York Knicks</td>
<td>32</td>
<td>50</td>
<td>8065</td>
<td>8298</td>
<td>0.97</td>
<td>0.406</td>
<td>0.39</td>
<td>0.016</td>
</tr>
<tr>
<td>2015-16</td>
<td>Oklahoma City Thunder</td>
<td>55</td>
<td>27</td>
<td>9038</td>
<td>8441</td>
<td>1.07</td>
<td>0.721</td>
<td>0.671</td>
<td>0.05</td>
</tr>
<tr>
<td>2015-16</td>
<td>Orlando Magic</td>
<td>35</td>
<td>47</td>
<td>8369</td>
<td>8502</td>
<td>0.98</td>
<td>0.445</td>
<td>0.427</td>
<td>0.018</td>
</tr>
<tr>
<td>2015-16</td>
<td>Philadelphia 76ers</td>
<td>10</td>
<td>72</td>
<td>7988</td>
<td>8827</td>
<td>0.9</td>
<td>0.200</td>
<td>0.122</td>
<td>0.078</td>
</tr>
<tr>
<td>2015-16</td>
<td>Phoenix Suns</td>
<td>23</td>
<td>59</td>
<td>8271</td>
<td>8817</td>
<td>0.94</td>
<td>0.291</td>
<td>0.28</td>
<td>0.011</td>
</tr>
<tr>
<td>2015-16</td>
<td>Portland Trail Blazers</td>
<td>44</td>
<td>38</td>
<td>8622</td>
<td>8554</td>
<td>1.01</td>
<td>0.528</td>
<td>0.537</td>
<td>0.009</td>
</tr>
</tbody>
</table>

For general queries, contact webmaster@press.princeton.edu
CHAPTER 1 APPENDIX: DATA TABLES

The Excel Data Table feature enables us to see how a formula changes as the values of one or two cells in a spreadsheet are modified. In this appendix we show how to use a one-way data table to determine how the accuracy of (2) for predicting team winning percentage depends on the value of exp. To illustrate let’s show how to use a one-way data table to determine how varying exp from 1 to 3 changes our average error in predicting an MLB’s team winning percentage (see Figure 1-2).

Step 1: We begin by entering the possible values of exp (1, 1.1, . . . , 3) in the cell range N7:N26. To enter these values we simply enter 1 in N7 and 1.1 in N8 and select the cell range N7:N8. Now we drag the cross in the lower right-hand corner of N8 down to N26.

Step 2: In cell O6 we enter the formula we want to loop through and calculate for different values of exp by entering the formula = J1. Then we select the “table range” N6:O26.

Step 3: Now we select Data Table from the What If section of the ribbon’s Data tab.

Step 4: We leave the row input cell portion of the dialog box blank but select cell G1 (which contains the value of exp) as the column input cell. After selecting OK we see the results shown in Figure 1-2. In effect, Excel has placed the values 1, 1.1, . . . , 3 into cell G1 and computed our MAD for each listed value of exp.
The letters t or f following a page number indicate a table or figure on that page.

Adler, Joseph, 65, 124, 127; *Baseball Hacks*, 65, 124, 127, 133
Albright, S. C., 120–121
Annis, David, 345
back-to-back games, NBA teams and, 538–540
*Baseball between the Numbers*, 67
*Baseball Hacks* (Adler), 54, 106
Bayes theorem 490–494; average and, 293–295; conjugate distributions and, 494–495; evaluating kickers and, 494–496; Monty Hall and, 492–493; posterior distribution and, 493–494; prior and, 497–498
Bellman, Richard, 198
Berk, Jonathan, 214
Berri, David (*Wages of Wins*), 164, 259
Berry, Scott, 534
Big Data Bowl, 230
Birnbaum, Phil, 183, 225
Boswell, Thomas, 24
Bowl Championship Series (BCS), 433; playoff alternatives and, 545; team rankings and, 433, 544–546
Braess paradox, 515
brainteaser, 150
Brodie, Mark, 117; *Every Shot Counts* and, 391
Brook, Stacey (*Wages of Wins*), 164, 259
Burke, Brian, 133–34; Bayesian evaluation of QBs and, 498; deferring in NFL OT and, 217; value of football states and, 174
bye week, NFL teams, 540–542
Cabot, Vic, 172
Calcutta auction, 474–476
calibration curve (referred to also as reliability or validation curve) 356–359, 389–390, 403, 459
Carter, Virgil, 172
clustering 335–336, 340, 372; k-means and, 513
collapses, probabilities and, 551; Mavericks and Raptors, 2019 and, 554; Mets, 1986 World Series and, 556–558; Patriots and Falcons, Super Bowl LI and, 552–554; Warriors and Cavaliers, 2016 NBA finals and, 551–552
conjugate distribution, 494
Index

conversions: the chart and, 198–203; dynamic programming and, 198; one-point versus two-point, 195–197
convex hull, 239–240, 322–323
Cook, Earnshaw (Percentage Baseball), 37
decision-making, baseball, 60, 70; base running and, 69–70; base stealing and, 67–69; bunting and, 67; expected value of random variables and, 63–65; experiments and random variables and, 63; possible states and, 61–62; runs per inning and, 72; tagging up from third base and, 71
decision-making, basketball, 341; corner three defense and, 335–339; end-game strategy and, 342–348; fouling with a three-point lead and, 344–348; lineup analysis and, 289–295; matchups and, 296–302; shooting three-point shot down two and, 342–344; two-person zero sum game theory (TPZSG), 337–338
decision-making, football, 178; accepting penalties and, 183; conversions and, 195–206; dynamic programming and, 198–203; field goal attempts and, 179–182; payoff matrix and, 186–193; possible states and, 170–174; punting and, 182–183; run-pass mix and, 184–185, 188–193; state values and, 175–177; two-person zero sum game theory (TPZSG) and, 186–193
Defense-Independent Component ERA (DICE), 53–55
dewan, John 76; Fielding Bible and, 76–79
Dolphin, Andrew, 90, 134
drafts: implied draft position value curve, NFL and, 222–226; NBA efficiency and, 305–306; NFL mock drafts and, 227; Winner’s Curse and, 227–228

eigenvalues and eigenvectors, 511–512, 514
e-sports 398–399; DOTA 2 hero adjusted +/− and, 401–403; DOTA 2 win probability and, 400–401; multiplayer online battle arena and, 399, 403; NBA2K league and, 404–405
Elo, 524; hyperparameters and, 525, 529–530; preseason NFL team ratings and, 526–528
Engelmann, Jeremias, 283, 287
fielders, evaluation of, 73; fielding percentage and, 73–75; pitch framing and, 83; range factor and, 75–76; runs saved, wins and, 77–78;
scores, runs and, 77; Statcast 78–79 and, 70; The shift and, 81–83
Four-Factor Model for NBA teams:
correlation lack and, 252–254; effective field goal percentage and, 249–250; free throw rate and, 251; rebound percentage and, 250; regression and, 254–256; success or failure of NBA teams and, 254; turnover percentage and, 250

Freakonomics (Levitt), 420–422

gambling: arbitrage betting and, 413–416; baseball and, 412–413; bettor biases and, 420; bettor profits and, 410–411; money line and, 411–12; money management and, 468–473; NBA win and gambling probabilities and, 451–453; NCAA tournament probabilities and, 453–457; NFL win and gambling probabilities and, 448–451; parlays and, 416–417; point spread setting system and, 424; strength of schedule and, 428, 519; team power ratings and, 423; teaser bets and, 417–419

game theory, 186–194, 335–337, 367–370

Glickman, Mark, 530; Glicko rating and, 530

Goldsberry, Kirk, 317, 324–325, 335, 507
golf, 391; average strokes gained and, 391–393; putting and, 391–393, 395; tee-to-green and, 393–395, 396–397

hockey: Corsi, 374–376; face-offs (skill-vs-luck) and, 380–381; Fenwick and, 375–376; power plays and, 382–384

Horowitz, Max, 171, 174, 483

Ilardi, Steve, 283

Laplace smoothing 437, 558
Lisk, Jason, 234

Ma, Jeff (The House Advantage), 348
matrix factorization 499–507; latent patterns and, 500–501; non-negative (NMF) and, 500–502; number of factors and, 504–505; shooting patterns 502–507
Maoussin, Michael (The Success Equation), 380
mediation variables (mediators), 236–237
multiplayer online battle arena game (MOBA), 399, 403

Nash equilibrium, 189, 337–340, 368, 516–518
natural experiment, 335

normal random variables, 107–109

officiating, NBA, 307–312; referee bias, testing for, 307–309; regression analysis and, 309–312; significant interaction and, 311–312
Oliver, Dean, 249
On-Base Percentage (OBP), 26
On-Base Plus Slugging (OPS), 26
optimization: regression and, 19–21;
regularization and, 273–275
ordinal logistic regression, 547–550
overtime, NFL, 211–212; fair solution
and, 213–215; mathematical model
of, 212–213
overtime college football, 207–210

p-values, 21–22
page rank, 519
Palmer, Pete, 24
park factors, 101–104; range factor
and, 76
payrolls: fair salary in MLB and,
98–99; NBA and, 303–304
Pelton, Kevin, 285
Percentage Baseball (Cook), 37
Perez, Tony, 127–131

pitchers, evaluation and forecasting
of: DICE and, 53–55; earned run
average (ERA) and, 44–45; ERA as
a predictive tool and, 47–49; MAD
and, 51; saves and, 46
pitchers, pitch count and, 133–136
platoon effect (splits), 67, 124–126
players in different eras, 139–141;
aging and, 534; all-time greats
and, 534–535; NBA player quality
changes and, 531–534
players value: adjusted +/- ratings,
NBA and, 267–275; ESPN RPM,
NBA and, 283, 283t; luck ad-
justed +/-, NBA and, 285; major
league equivalents and, 151–153;
Marcel projection and, 153–156;
NBA draft efficiency and, 305–306;
NBA efficiency rating and, 260–261,
261t; NFL draft picks and, 222–227;
plate appearances and, 82–83;
PIMP, NBA and, 285; Player Ef-
ficiency Rating (PER), NBA and,
261; pure +/- ratings, NBA and,
265–266; RAPTOR, NBA and
285, 283t; salaries and, 98–99,
303–305; Win Scores and Wins
Produced 283t; Winner’s Curse
and, 227–228; win scores and wins
produced, NBA and, 262–264

player tracking data, 229, 321; convex
hull and, 239–240, 322–323;
corner 3s and, 329–340; defense
and, 323–325; expected point value
(DeepHoops) and, 325–329; NBA
and, 321–340; NFL and, 229–246;
offensive line and, 239–242; pass-
ing (NFL) and, 230–239, 242–246
Player Win Averages, 84; baserunning
and, 89; fielding ratings and, 88;
Win Expectancy Finder and, 84–85
Pluto, Terry (Tall Tales), 265
Poisson random variables, 143–145,
180, 358, 377–379, 442–445
Pomeroy, Ken, 350
predictions in sports, 57–58; Bradley-
Terry model and, 387–390; ERA
using DICE and, 47, 54; gener-
alized linear models and, 174,
180–181, 355–359; least squares
team ratings and, 424–428; Marcel
projection and, 153–156; NCAA
tournament and, 453–457; net-
work ratings and, 518–523; NFL
playoffs and, 457–458; probability
calibration (Platt scaling) and,
458–462; Pythagorean theorems
and, 8–10, 385; ridge regression
and, 273–275; soccer and, 442–446
Price, Joseph, 307, 312
principal component analysis, 513–514
probability matching, 370
probability theory, 63; consecutive
no-hitters and, 148–150; DiMag-gio’s 56-game hitting streak and, 146–148; expected value and, 63–65; experiments and random variables and, 63; independent events and, 144; law of conditional expectation and, 65; law of rare events and, 143; NBA win and gambling probabilities and, 293–94; NCAA tournaments and, 294–96; NFL win and gambling probabilities and, 290–92; perfect games and, 144–145; pitching consecutive no hitters and, 148–150; sports collapses and, 551–558; team wins, betting and, 447–462.
Pythagorean Theorem: baseball and, 3–7; football and basketball and, 8–10; forecasting and, 7–8; mean absolute deviation (MAD) and, 5–6, 5f, 6f; runs, wins and, 78; Trout and, 40–41; volleyball and, 385
quarterbacks, rating of, 164–169; Burke’s regression and, 166–169, 166f, 167f
random variables, 63–65; expected value and, 64; independence and, 114, 144; normal distribution and, 107–109; Poisson and, 143–144; variances and, 291–292.
ranking teams: BCS and, 433–439; Elo and, 524–525; evaluating offenses and defenses and, 431–433; Glicko and, 530; home edge and, 290, 388, 424–425, 447, 524; mean absolute errors and, 428–431; multiplicative ratings and, 439–444; power ratings and, 424, 447; strength of schedule and, 428, 519; wins and losses and, 433–439
rare events, probability of, 143
Reese, S., 534
regressions, 18–20; Excel and, 29–30; logistic and, 179–182, 216–218, 218t, 233–235, 233t, 234f, 235t, 241, 333, 334t, 356–357, 357t; NBA officiating and, 309–312; Oliver’s Four Factors and, 255; ordinal logistic and, 547–550, 549t; p-values and, 21; quarterback ratings and, 166–169; ridge and, 273–275; scoring margins, NFL teams and, 159–162; Skellam, soccer and, 445
Romer, David, 172–174
Rosen, Peter A., 207
R-squared value (RSQ), 48–51; Excel and, 56
Runs Created formula, 13–15, 22, 152
sabermetrics, 3, 80, 380
Sagarin, Jeff, 31, 37; clutch hitters and, 129; expected values (NFL) and, 172–173, 173t, NFL overtime, and, 219–221; team ratings and, 431, 438–439, 447–448, 453–454, 539–542; WINVAL, 268–273
Schatz, Aaron, 58
Schmidt, Martin (Wages of Wins), 164, 259
Shandler, Ron, 58
Silver, Nate (The Signal and the Noise), 497
skill curves, 515–516
slugging percentage (SLG), 12, 26–27, 27f, 152
Smith, Ozzie, 75
Sobel test (mediation variable), 237
Spurr, David, 45
Stoll, Greg, Win Expectancy Finder and, 84
streakiness: baseball hitting and, 120–122; hot hand and, 107, 113–122; hot teams and, 122–123; hypothesis testing and, 112–113; normal random variables and, 107–109; random sequences and, 105–106; Wald Wolfowitz runs test (WWRT) and, 107, 111–112; z-scores and, 109–111
synergy: isolation and, 319–320; pick-and-roll and, 313–317; play type data, 313; post-ups and, 317–319

**Tall Tales** (Pluto), 265
Tango, Tom: Marcel projection system, 153–156; *The Book* and, 104, 134; weighted on-base average (wOBA) and, 93–94
Thaler, Richard, 222, 305
Thomson, Bobby, 84, 91
Tversky, A., 106–107, 113
two proportion z-test, 383–384

Ulam, Stanislaw, 32

<table>
<thead>
<tr>
<th>Name</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vallone, R.</td>
<td>107, 113–114</td>
</tr>
<tr>
<td>Ventura, Sam</td>
<td>174, 373, 380, 483</td>
</tr>
<tr>
<td>volleyball: Bradley-Terry model and</td>
<td>387–390; Pythagorean and, 385</td>
</tr>
<tr>
<td>Vollman, Rob</td>
<td><em>(Stat Shot)</em>, 374</td>
</tr>
<tr>
<td>Wages of Wins</td>
<td><em>(Berri, Schmidt, Brook)</em>, 164, 259</td>
</tr>
<tr>
<td>Wald Wolfowitz runs test (WWRT)</td>
<td>107, 111</td>
</tr>
<tr>
<td>Wilson, Rick L.</td>
<td>207–208</td>
</tr>
<tr>
<td>Win Expectancy Finder</td>
<td>84, 85f, 556, 557f</td>
</tr>
<tr>
<td>Winner’s Curse: MLB free agents and</td>
<td>99; NFL draft and, 227–228</td>
</tr>
<tr>
<td>win probability: added baseball and</td>
<td>84–88; DOTA 2 and, 399–404, 401t, 403f; in-game and, 553–554, 555–556; NFL and, 171, 174–175, 204–206, 218t; PageRank and, 521; Platt scaling and, 458–462; pulling the goalie and, 376–379; team ratings and, 447–458; volleyball and, 387–390, 390f</td>
</tr>
<tr>
<td>WINVAL ratings</td>
<td>268; impact ratings and, 277–278; Khris Middleton contract and, 279; top 10 players and, 271f</td>
</tr>
<tr>
<td>Wolfers, Justin</td>
<td>307, 312</td>
</tr>
<tr>
<td>Woolner, Keith</td>
<td>135</td>
</tr>
<tr>
<td>Yurko, Ron</td>
<td>171, 174, 483</td>
</tr>
</tbody>
</table>

z-scores, 109–111, 113