****

9B18E015

The Best Hitter in Baseball

Kyle Maclean wrote this case solely to provide material for class discussion. The author does not intend to illustrate either effective or ineffective handling of a managerial situation. The author may have disguised certain names and other identifying information to protect confidentiality.

*This publication may not be transmitted, photocopied, digitized or otherwise reproduced in any form or by any means without the permission of the copyright holder. Reproduction of this material is not covered under authorization by any reproduction rights organization. To order copies or request permission to reproduce materials, contact Ivey Publishing, Ivey Business School, Western University, London, Ontario, Canada, N6G 0N1; (t) 519.661.3208; (e)* [*cases@ivey.ca*](mailto:cases@ivey.ca)*;* [*www.iveycases.com*](http://www.iveycases.com)*.*

Copyright © 2018, Ivey Business School Foundation Version: 2018-10-04

THE LIST

Kevin DeKuyper, a sports journalist based in Houston, Texas, was scratching his head in confusion at his editor’s latest idea. His editor had emailed him the night before proposing an article on the best (and worst) Major League Baseball (MLB) hitters of all time. The first step was to provide a list of the players who would be profiled. This meant finding the 10 best (and worst) batters. Luckily, DeKuyper editor had included helpful career batting statistics on which to base the analysis (see Exhibit 1).

The file contained the name of each player (excluding pitchers) from 1871 to 2016, the players’ number of at-bats (batting attempts), the number of successful hits for each player, and using the two data points, the players’ batting average. The batting average was calculated as the number of hits divided by the number of at-bats. For example, a 0.300 batting average (commonly referred to as “batting 300”) was considered good, and a higher batting average meant a better player.

In his small desk cubicle, DeKuyper looked at the list of over 9,000 names in the Microsoft Excel file. The email ended with a note from his editor with a suggestion: “Maybe just rank them based on batting average and take the top and bottom 10 hitters. That should be sufficient.” He hoped this would be true but wondered if that was the best approach. DeKuyper remembered a recent article about Bayesian updating and beta distribution (see Exhibit 2) and wondered if this could help him quantify uncertainty in hitting rates. For example, after ranking the players could he quantify the probability that he ranked them incorrectly? Being a data-based article meant that he needed to justify his choices, and he wanted his decisions to be based on sound statistical grounds.

Exhibit 1: Sample Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Player Name | Hits | At-Bats | Batting Average |
| 1 | Hank Aaron | 3,771 | 12,364 | 0.304998 |
| 2 | Tommie Aaron | 216 | 944 | 0.228814 |
| 3 | Andy Abad | 2 | 21 | 0.095238 |
| 4 | John Abadie | 11 | 49 | 0.22449 |
| 5 | Ed Abbaticchio | 772 | 3,044 | 0.253614 |
| 6 | Fred Abbott | 107 | 513 | 0.208577 |
| 7 | Jeff Abbott | 157 | 596 | 0.263423 |
| 8 | Kurt Abbott | 523 | 2,044 | 0.255871 |
| 9 | Ody Abbott | 13 | 70 | 0.185714 |
| 10 | Frank Abercrombie | 0 | 4 | 0 |
|  |  |  |  |  |
| 9,503 | Jon Zuber | 34 | 136 | 0.25 |
| 9,504 | Julio Zuleta | 43 | 174 | 0.247126 |
| 9,505 | Mike Zunino | 219 | 1,125 | 0.194667 |
| 9,506 | Bob Zupcic | 199 | 795 | 0.250314 |
| 9,507 | Frank Zupo | 3 | 18 | 0.166667 |
| 9,508 | Paul Zuvella | 109 | 491 | 0.221996 |
| 9,509 | Dutch Zwilling | 364 | 1,280 | 0.284375 |

Source: Created by the case author using data from “Download Lahman’s Baseball Database,” SeanLahman.com: Baseball, Data, and Storytelling, accessed August 27, 2018, www.seanlahman.com/baseball-archive/statistics.

Exhibit 2: BETA DISTRIBUTION

The beta distribution is a continuous probability distribution that takes two parameters ( and ), both of which must be strictly positive. The beta distribution is commonly used to model parameter uncertainty because it is bounded between 0 and 1 and is well suited to describe the uncertainty about the probability of success. The formal definition of the distribution is given as follows:

|  |  |
| --- | --- |
| **Mean** |  |
| **Variance** |  |

The beta distribution’s definition of its probability density function and associated cumulative density function is complicated but available online.\*

The beta distribution also has “nice” properties when performing Bayesian updating. Specifically, it has the following *updating rule*. Suppose we assume that the binomial distribution parameter *p* is unknown and distributed according to a beta distribution with parameters alpha and beta (). Suppose we observe *X* successes in *N* trials as an outcome. The posterior (or updated) distribution of *p* will *also* be beta distributed with parameters ().

In Microsoft Excel, the function “BETA.DIST(*X*, Alpha, Beta, Cumulative?)” returns the probability density function (*pdf*) or cumulative distribution function (*cdf*). The first argument to the function is the value at which to generate the probability (or cumulative) density function. The second and third arguments are the alpha and beta parameters. The fourth parameter, if set to TRUE, returns the *cdf*; if set to FALSE, it returns the *pdf*. The function has optional fifth and sixth parameters, but these are not necessary for the most common use of the function.

The Excel function “BETA.INV(*X*, Alpha, Beta)” returns the value *probability* such that BETA.DIST (probability, alpha, beta, TRUE) = *X*. The first argument to the function is the value at which to evaluate the function. The second and third arguments are the alpha and beta parameters. The function has optional fourth and fifth parameters, but these are not necessary for the most common use of the function.

\*For example, see “Beta Distribution,” Wolfram MathWorld, accessed August 27, 2018, http://mathworld.wolfram.com/BetaDistribution.html; “1.3.6.6.17. Beta Distribution,” Engineering Statistics Handbook, accessed August 27, 2018, www.itl.nist.gov/div898/handbook/eda/section3/eda366h.htm.

Source: “BETA.DIST Function,” Microsoft, accessed August 7, 2018, https://support.office.com/en-us/article/beta-dist-function-11188c9c-780a-42c7-ba43-9ecb5a878d31; “BETA.INV Function,” Microsoft, accessed August 7, 2018, https://support.office.com/en-us/article/beta-inv-function-e84cb8aa-8df0-4cf6-9892-83a341d252eb.