Baseball Analytics Final Report

Ronald Michaels

Spring 2020 Semester

This paper will describe my pesonal project for the independent study of Baseball Analytics at the University of Rochester. The team of students consisted of Maya Haigis, Luke Gerstner, John Polimeni, and myself. The academic advisors watching over our work were Professor Anand, Professor Mort, and Head Baseball Coach Joe Reina.

My personal project for this final report is based on comparing the predictions Coach Reina had for our offensive player's versus what the computer model predicts. In terms of desired statistics, Coach Reina and I decided that the four main statistics we wanted to highlight were: Batting Average (AVG), On Base Percentage (OBP), Slugging Percentage (SLG), and On Base Plus Slugging (OPS). **First-year player's and sophomores who were red-shirted their first-year will not be included in this report due to their limited exposure to college baseball at this time.**

The following table displays the predictions Coach Reina had for our elligible offensive player's

```
##         Player.Name   AVG   OBP  SLG   OPS Class.Year
## 1       Aaron Craig 0.290 0.360 0.35 0.710     Senior
## 2       David Rieth 0.410 0.420 0.60 1.020     Senior
## 3        Jake Hertz 0.335 0.400 0.40 0.800     Senior
## 4     Kyle Trombley 0.250 0.275 0.30 0.575     Senior
## 5     Ryland McNabb 0.250 0.333 0.25 0.583     Senior
## 6        Zach Miraz 0.310 0.400 0.45 0.850     Senior
## 7  Andrew Bankovich 0.189 0.200 0.35 0.550     Junior
## 8   Steve Pickering 0.265 0.375 0.38 0.755     Junior
## 9    Brian McKinsey 0.320 0.400 0.41 0.810  Sophomore
## 10        Harper Sy 0.310 0.450 0.40 0.850  Sophomore
## 11     Jacob Matzat 0.320 0.350 0.42 0.770  Sophomore
## 12    Joseph Rende  0.350 0.460 0.45 0.910  Sophomore
## 13     Luke Piontek 0.280 0.333 0.40 0.733  Sophomore
```

The next step in this process is to input average offensive statistics from not just our home conference of the Liberty League, but of all division three baseball teams on the NCAA website. Due to the untimely cancellation of our baseball season as a result of COVID-19, the statistics will be weighted based on the player's statistics over time. For example, current seniors will be predicted based on statistics from the years 2017-2020, juniors from the years 2018-2020, and sophomores from the years 2019-2020.

Comparison Statistics from 2017-2020:

2017 Liberty League Baseball Average Statisitcs

```
##          AVG       OBP       SLG       OPS
## 1: 0.2793333 0.3633333 0.3767778 0.7401111
```

2018 Liberty League Baseball Average Statisitcs

```
##          AVG       OBP       SLG   OPS
## 1: 0.2818889 0.3694444 0.3845556 0.754
```

2019 Liberty League Baseball Average Statisitcs

```
##           AVG       OBP       SLG       OPS
## 1: 0.2831111 0.3727778 0.3855556 0.7583333
```

2020 Liberty League Baseball Average Statisitcs

```
##       AVG    OBP    SLG    OPS
## 1: 0.3026 0.3883 0.4309 0.8192
```

2017 NCAA Average Statistics

```
##           AVG       OBP       SLG       OPS
## 1: 0.2884667 0.3743111 0.4653711 0.8396822
```

2018 NCAA Average Statistics

```
##           AVG       OBP       SLG       OPS
## 1: 0.2875227 0.3776364 0.4686255 0.8462619
```

2019 NCAA Average Statistics

```
##           AVG       OBP       SLG       OPS
## 1: 0.2856222 0.3767333 0.4655861 0.8423195
```

2020 NCAA Average Statistics

```
##           AVG       OBP       SLG       OPS
## 1: 0.2821111 0.3801778 0.4789247 0.8591025
```

Each data frame will be weighted based on the player's year in school. In my estimate, it is important to value the player's earlier statistics versus their later statistics due to their perceived growth over time. For example, if a player were to have a batting average of 0.300 their first year versus a batting average of 0.300 their senior year, I am choosing to weight their first year statistics more than their senior year statistics. The weight for each class year will follow this scale:

For Current Seniors:
First Year Statistics: .40 = 40%
Sophomore Year Statistics: .30 = 30%
Junior Year Statistics: .20 = 20%
Senior Year Statistics: .10 = 10%

For Current Juniors:
First Year Statistics: .50 = 50%
Sophomore Year Statistics: .35 = 35%
Junior Year Statistics: .15 = 15%

For Current Sophomores:
First Year Statistics: .85 = 85%
Sophomore Year Statistcs: .15 = 15%

The next step in the prediction process is to see how these player's compared to their peers (Liberty League + Overall NCAA). In addition, their percent difference from these peers will be calculated as well.

Senior Mean Batting Average (AVG) - Weighted = 0.2850806
Junior Mean Batting Average (AVG) - Weighted = 0.2857346
Sophomore Mean Batting Average (AVG) - Weighted = 0.285565

Senior Mean On Base Percentage (OBP) - Weighted = 0.372966
Junior Mean On Base Percentage (OBP) - Weighted = 0.3755705
Sophomore Mean On Base Percentage (OBP) - Weighted = 0.3761781

Senior Mean Slugging Percentage (SLG) - Weighted = 0.4270123
Junior Mean Slugging Percentage (SLG) - Weighted = 0.4304819
Sophomore Mean Slugging Percentage (SLG) - Weighted = 0.4299721

Senior Mean On Base Plus Slugging Percentage (OPS) - Weighted = 0.7999783
Junior Mean On Base Plus Slugging Percentage (OPS) - Weighted = 0.8060524
Sophomore Mean On Base Plus Slugging Percentage (OPS) - Weighted = 0.8061501

Average Calculation with Weighted Statistics per Player

```
##          Player.Name   AVG   OBP   SLG   OPS Class.Year
##  1:      Aaron Craig 0.221 0.306 0.241 0.547     Senior
##  2:      David Rieth 0.338 0.429 0.460 0.889     Senior
##  3:       Jake Hertz 0.342 0.389 0.414 0.803     Senior
##  4:    Kyle Trombley 0.222 0.296 0.284 0.580     Senior
##  5:    Ryland Mcnabb 0.263 0.422 0.289 0.711     Senior
##  6:       Zach Miraz 0.219 0.334 0.248 0.582     Senior
##  7: Andrew Bankovich 0.258 0.323 0.384 0.707     Junior
##  8:  Steve Pickering 0.326 0.353 0.391 0.744     Junior
##  9:   Brian McKinsey 0.186 0.326 0.243 0.569  Sophomore
## 10:        Harper Sy 0.300 0.335 0.383 0.718  Sophomore
## 11:     Jacob Matzat 0.298 0.380 0.352 0.732  Sophomore
## 12:     Joseph Rende 0.281 0.353 0.344 0.696  Sophomore
## 13:     Luke Piontek 0.094 0.292 0.124 0.356  Sophomore
```

Percent Difference With Weighted Averages versus National Averages

```
##          Player.Name    AVG    OBP    SLG    OPS Class.Year
##  1:      Aaron Craig -0.225 -0.180 -0.436 -0.316     Senior
##  2:      David Rieth  0.186  0.150  0.077  0.111     Senior
##  3:       Jake Hertz  0.200  0.043 -0.030  0.004     Senior
##  4:    Kyle Trombley -0.221 -0.206 -0.335 -0.275     Senior
##  5:    Ryland Mcnabb -0.077  0.131 -0.323 -0.111     Senior
##  6:       Zach Miraz -0.232 -0.104 -0.419 -0.272     Senior
##  7: Andrew Bankovich -0.097 -0.140 -0.108 -0.123     Junior
##  8:  Steve Pickering  0.141 -0.060 -0.092 -0.077     Junior
##  9:   Brian McKinsey -0.349 -0.133 -0.435 -0.294  Sophomore
## 10:        Harper Sy  0.051 -0.109 -0.109 -0.109  Sophomore
## 11:     Jacob Matzat  0.044  0.010 -0.181 -0.092  Sophomore
## 12:     Joseph Rende -0.016 -0.062 -0.200 -0.137  Sophomore
## 13:     Luke Piontek -0.671 -0.224 -0.712 -0.558  Sophomore
```

Now that the weighted averages are calculated, the multiple linear regression that will be performed will be for the four main statistics (AVG,OBP,SLG,OPS) based on class year and average at bats per game.

Number Of At Bats Per Game Over Career

```
##          Player.Name Average.AB.Per.Game Class.Year
##  1:      Aaron Craig           1.5042735     Senior
##  2:      David Rieth           3.4102564     Senior
##  3:       Jake Hertz           2.6153846     Senior
##  4:    Kyle Trombley           2.6068376     Senior
##  5:    Ryland Mcnabb           0.4444444     Senior
##  6:       Zach Miraz           0.7350427     Senior
##  7: Andrew Bankovich           1.7241379     Junior
##  8:  Steve Pickering           3.3333333     Junior
```

```
##  9:   Brian McKinsey           1.1568627  Sophomore
## 10:        Harper Sy           2.4117647  Sophomore
## 11:     Jacob Matzat           2.2352941  Sophomore
## 12:     Joseph Rende           2.1568627  Sophomore
## 13:     Luke Piontek           0.2549020  Sophomore
```

Regression Analysis

```
##          Player.Name LR.AVG LR.OBP LR.SLG LR.OPS Class.Year
##  1:      Aaron Craig  0.248  0.357  0.296  0.651     Senior
##  2:      David Rieth  0.346  0.384  0.429  0.824     Senior
##  3:       Jake Hertz  0.305  0.373  0.374  0.752     Senior
##  4:    Kyle Trombley  0.304  0.373  0.373  0.751     Senior
##  5:    Ryland Mcnabb  0.194  0.342  0.222  0.554     Senior
##  6:       Zach Miraz  0.208  0.346  0.242  0.580     Senior
##  7: Andrew Bankovich  0.251  0.327  0.331  0.652     Junior
##  8:  Steve Pickering  0.333  0.349  0.444  0.799     Junior
##  9:   Brian McKinsey  0.207  0.330  0.255  0.570  Sophomore
## 10:        Harper Sy  0.271  0.348  0.343  0.684  Sophomore
## 11:     Jacob Matzat  0.262  0.346  0.331  0.668  Sophomore
## 12:     Joseph Rende  0.258  0.345  0.325  0.661  Sophomore
## 13:     Luke Piontek  0.161  0.317  0.192  0.488  Sophomore
```

Multiply the predicted values from the above regression table by their specific percent differences to obtain the final computer model outputs.
Of Note: OPS does not precisely equal OBP + SLG due to rounding errors as a result of each statistic being held to three decimal places.

```
##          Player.Name   AVG   OBP   SLG   OPS Class.Year
##  1:      Aaron Craig 0.192 0.293 0.167 0.445     Senior
##  2:      David Rieth 0.410 0.442 0.462 0.915     Senior
##  3:       Jake Hertz 0.366 0.389 0.363 0.755     Senior
##  4:    Kyle Trombley 0.237 0.296 0.248 0.544     Senior
##  5:    Ryland Mcnabb 0.179 0.387 0.150 0.493     Senior
##  6:       Zach Miraz 0.160 0.310 0.141 0.422     Senior
##  7: Andrew Bankovich 0.227 0.281 0.295 0.572     Junior
##  8:  Steve Pickering 0.380 0.328 0.403 0.737     Junior
##  9:   Brian McKinsey 0.135 0.286 0.144 0.402  Sophomore
## 10:        Harper Sy 0.285 0.310 0.306 0.609  Sophomore
## 11:     Jacob Matzat 0.274 0.349 0.271 0.607  Sophomore
## 12:     Joseph Rende 0.254 0.324 0.260 0.570  Sophomore
## 13:     Luke Piontek 0.053 0.246 0.055 0.216  Sophomore
```

The final portion of this report is comparing these computer results to Coach Reina's predictions.
Percent Difference versus Coach Reina
Positive percentages mean Coach Reina's predictions are above the computer generated outcomes by the given percentage.
Negative percentages mean Coach Reina's predictions are below the computer generated outcomes by the given percentage.

```
##          Player.Name    AVG    OBP    SLG    OPS Class.Year
##  1:      Aaron Craig    51%  22.9% 109.6% 109.6%     Senior
##  2:      David Rieth     0%    -5%  29.9%  29.9%     Senior
##  3:       Jake Hertz  -8.5%   2.8%  10.2%  10.2%     Senior
```

```
##  4:    Kyle Trombley   5.5%  -7.1%    21%    21%     Senior
##  5:    Ryland Mcnabb  39.7%   -14%  66.7%  66.7%     Senior
##  6:       Zach Miraz  93.8%    29% 219.1% 219.1%     Senior
##  7: Andrew Bankovich -16.7% -28.8%  18.6%  18.6%     Junior
##  8:  Steve Pickering -30.3%  14.3%  -5.7%  -5.7%     Junior
##  9:   Brian McKinsey   137%  39.9% 184.7% 184.7%  Sophomore
## 10:        Harper Sy   8.8%  45.2%  30.7%  30.7%  Sophomore
## 11:     Jacob Matzat  16.8%   0.3%    55%    55%  Sophomore
## 12:     Joseph Rende  37.8%    42%  73.1%  73.1%  Sophomore
## 13:     Luke Piontek 428.3%  35.4% 627.3% 627.3%  Sophomore
```

After speaking to Coach Reina about the results, I have a newfound appreciation for what a coach takes into consideration versus what a computer takes in. For example, a coach takes into account the mental perception of pre-season performance, growth in the weight room, and intellectual maturity. The computer, on the other hand, only takes in the desired concrete information. Therefore, my conclusion for this specific project is that the optimal solution for perfecting predictions is to not only analyze a player's raw statistics, but to also look at intangible traits that are blind to machines. It is the combination of what coaches describe as "the eye test" and what analysts call "raw data" to create peak analysis.