

Baseball Hall of Fame: Examining Factors for Enshrinement

Maria McMahon, Courtney Taylor, and Jesus Riera Sanchez

Department of Mathematics and Statistics

Faculty Mentors: Dr. David Agard and Dr. Joseph Nolan



INTRODUCTION

The final recognition for professional baseball players deemed to have had outstanding careers is enshrinement in the Baseball Hall of Fame¹ (a museum in Cooperstown, NY). This research project examines the career records for enshrined players and investigates the impact of performance variables on the chance of enshrinement. Career statistics for all HOF-eligible retired players are used to:

- Describe career progression for Hall of Famers versus non Hall of Famers. (based on all eligible retired players)
- Investigate the effect of position (e.g. catcher) on the chance of enshrinement.
- Identify variables that significantly affect the chance of enshrinement into the Hall of Fame.
- Use a predictive model to assess the chances of enshrinement for all players not retired by 2010.

DATA

Minimum eligibility for the Hall of Fame is a ten year career and retirement of at least five years. Each year a screening committee establishes the ballot of players for voting. Annual statistics for all eligible players retired by 2010 were obtained from Lahman's Baseball Database². The following variables were collected based on primary position:

- Batters: hits, homeruns, runs batted in (RBI), and batting average.
- Pitchers: wins, saves, strike outs, and earned run average (ERA).

METHODS

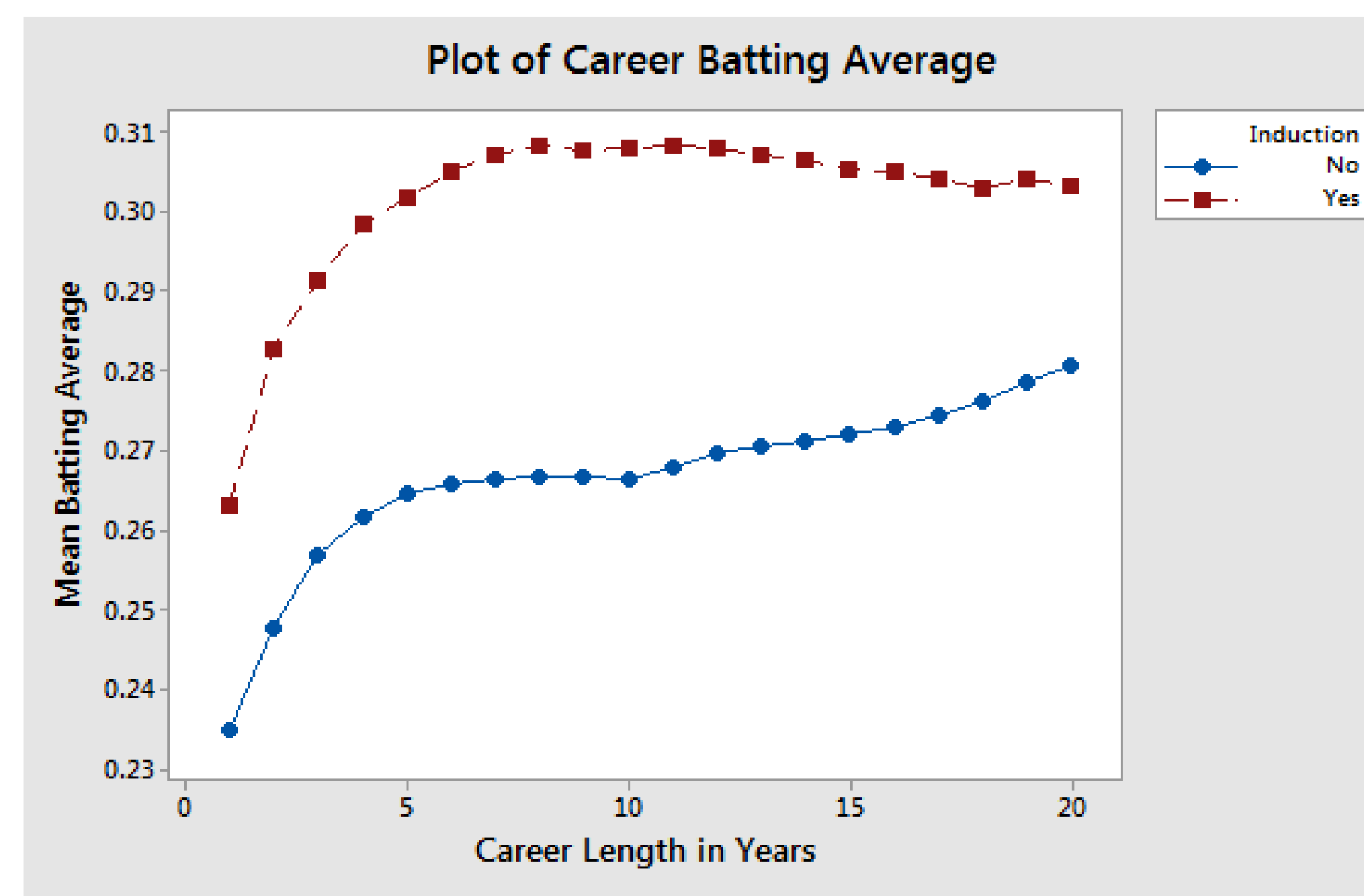
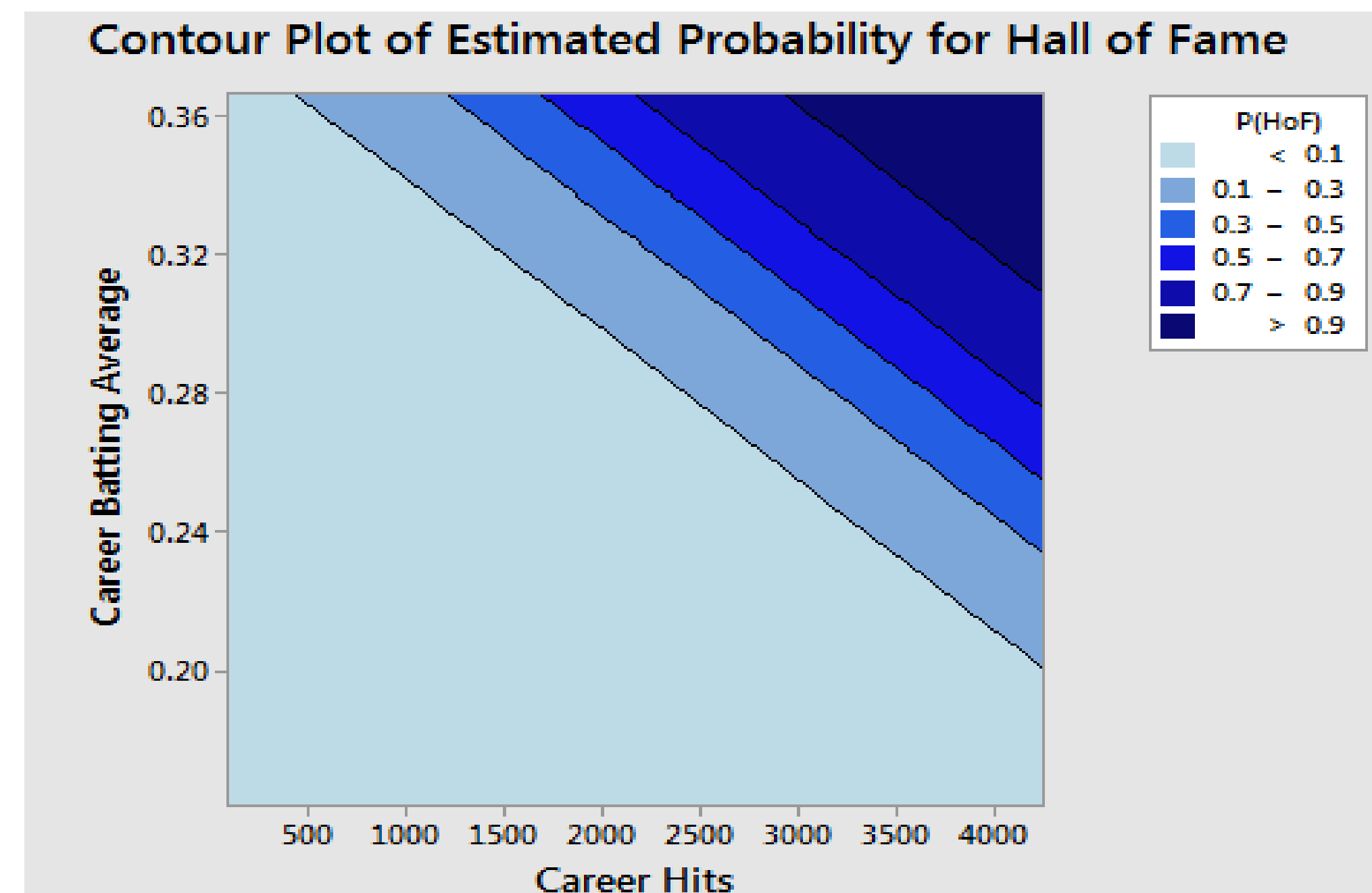
The Statistical Analysis System (SAS)[®] (Version 9.3, copyright 2010) was employed to create partial and final career totals for all variables. Graphical descriptive techniques highlighted differences in career progression between Hall of Famers and non Hall of Famers. Logistic regression analysis identified significant variables in estimating the likelihood of enshrinement in the Hall of Fame. For current players career end statistics were projected from partial career statistics. Logistic models were then used to predict the probability of enshrinement for selected current players.

RESULTS

Hitting

Logistic regression results found that hits, batting average, runs batted in, and position played are important predictors of induction into the Hall of Fame.

Variables	P-Value
Hits	<0.0005
Home Runs	0.113
RBI	0.003
Batting Average	<0.0005
Position	<0.0005



Top 10 Most Likely Future Batting Candidates

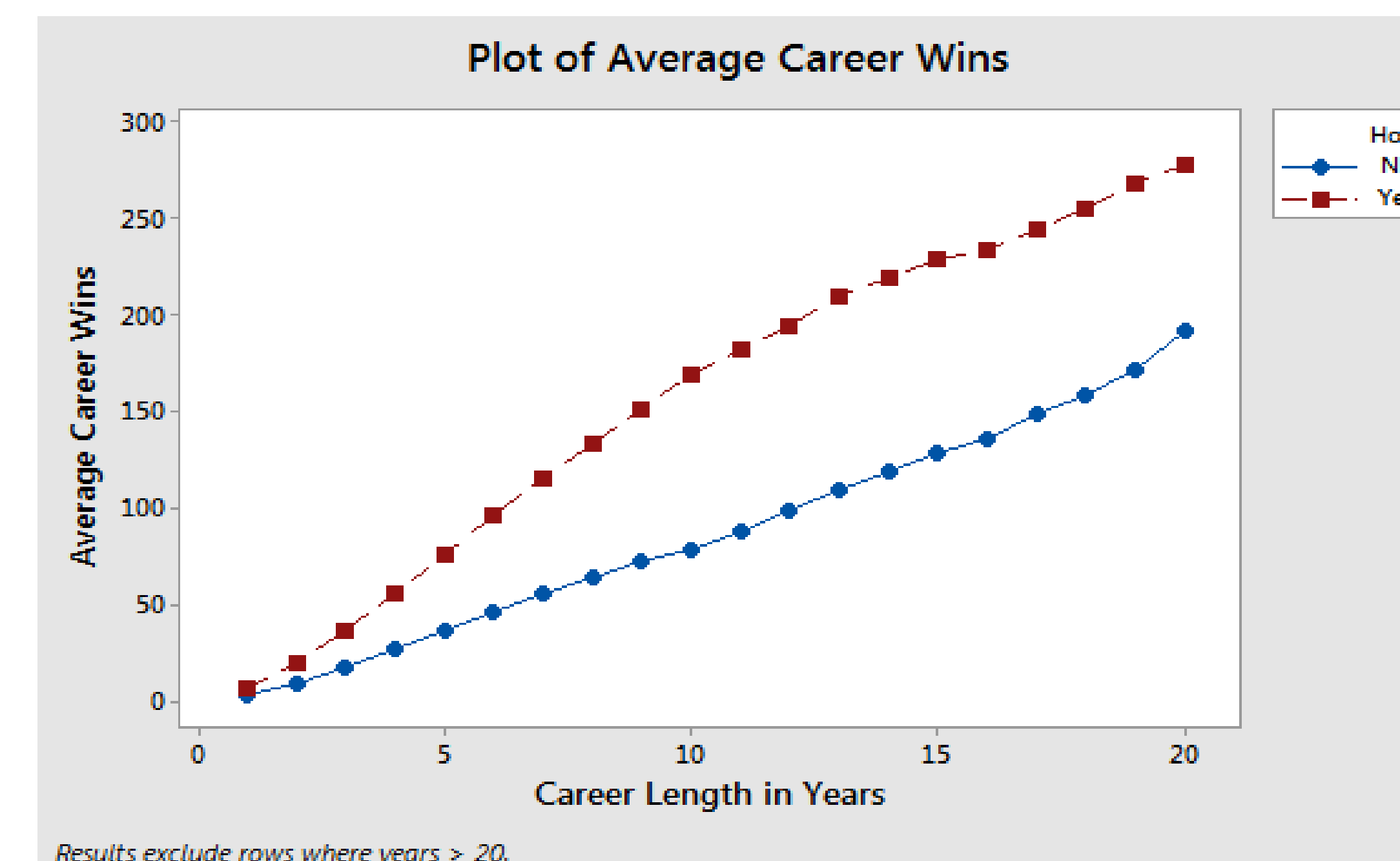
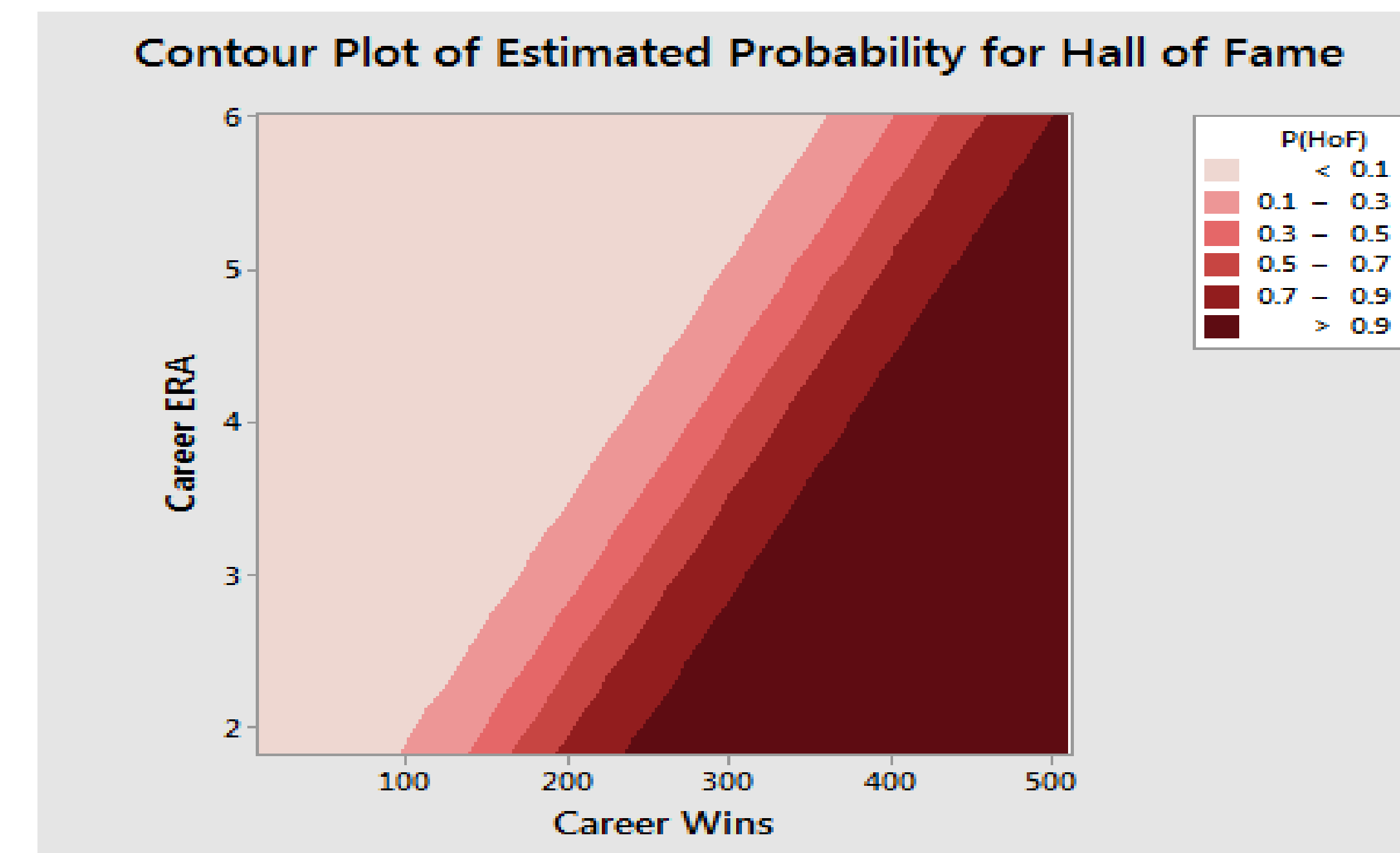
Name	P(HoF)
1 Alex Rodriguez, SS	94.7%
2 Derek Jeter, SS (R)	94.6%
3 Manny Ramirez, OF (R)	89.8%
4 Albert Pujols, 1B	89.1%
5 Vladimir Guerrero, OF (R)	86.1%
6 Chipper Jones, 3B (R)	84.5%
7 Todd Helton, 1B (R)	81.2%
8 Miguel Cabrera, 1B	79.8%
9 Ivan Rodriguez, C (R)	74.8%
10 Ichiro Suzuki, OF	71.8%
11 Adrian Beltre, 3B	66.1%
12 Bobby Abreu, OF (R)	54.2%
13 Magglio Ordonez, OF (R)	52.5%
14 Johnny Damon, OF (R)	51.1%
15 David Ortiz, DH	50.9%

R ~ Retired as of 2011

Pitching

Logistic regression results found that wins, saves, strike outs, and ERA are significant in the prediction of induction into the Hall of Fame.

Variables	P-Value
Wins	<0.0005
ERA	<0.0005
Saves	<0.0005
Strike Outs	0.083



Top 10 Most Likely Future Pitching Candidates

Name	P(HoF)
1 Clayton Kershaw	96.9%
2 Mariano Rivera* (R)	95.3%
3 Craig Kimbrel*	94.6%
4 Felix Hernandez	93.0%
5 Francisco Rodriguez*	74.6%
6 CC Sabathia	69.4%
7 Justin Verlander	66.4%
8 Jonathan Papelbon*	57.6%
9 Huston Street*	48.6%
10 Zack Greinke	45.6%
11 Madison Bumgarner	43.1%
12 Jon Lester	37.8%
13 Cole Hanel	36.3%
14 Aroldis Chapman*	35.4%
15 Mark Buehrle	34.1%

* ~ Relief Pitcher

CONCLUSIONS

- The logistic regression analysis correctly “fit” 95% of eligible retired players.
- For hitters important variables include hits, batting average, RBI, and position.
- For pitchers important variables include strike outs, saves, wins, and ERA.
- The predictive models seem to show a tendency toward hitters in the estimated chance of enshrinement. The model finds only six current pitchers with a greater than fifty percent chance of enshrinement, while there are fifteen such current hitters.

FUTURE

Possibilities for further studies include:

- Incorporation of more advanced performance metrics (such as wins above replacement, on base percentage, and WHIP) in the logistic models.
- Improved analysis on the effect of the time frame played. For example using year to year variability.
- Incorporation of the effect of position played in the predictive models.
- Study interaction between homeruns and “Steroid Era” on HOF selection.

REFERENCES

- ¹ Retrieved June 13, 2016 from <http://www.baseballhall.org>
- ² Retrieved June 13, 2016 from <http://www.SeanLahman.com>

ACKNOWLEDGEMENT

This research was a product of the UR-STEM summer undergraduate research experience supported by NKU *FORCE: Focus on Occupations, Recruiting, Community, and Engagement*.