

Predicting Post-Peak MLB Player Performance Per Position

Sammy Alali, Joe Datz, Henry Gise, Brendon Gu, Dexter Harrell, Juliette Wong, Jeffrey Wheeler (Faculty)

Math 1103 - BIG Problems, Dept. of Mathematics, University of Pittsburgh

Problem Statement

Given a player's past performance and playing history, can we create a predictive model that assesses the ability of a position player to play each defensive position?



The Plaid Panthers:

The **Plaid Panthers** is a professional baseball team based in downtown Philadelphia, PA that is a part of the Imaginary Baseball League (IBL).

- Founded by Big ProblemsTM
- Big-Market Team, hoping to use data analytics to help cut down their costs and continue winning
- Historically always win their division, mainly due to having more money. However, with more small teams using data analytics to improve themselves, they want to use analytics to stay on top.
- Not struggling financially, trying to go under the luxury tax threshold
- Preference towards spending a lot of money, but only if it is a good return on investment

	Team	W	L	%	GB
	Plaid Panthers	103	59	.636	-
	Houston Trashtros	101	61	.623	2
	Tampa Bay Yankers	96	66	.593	5
	Bigelow Bricklayers	69	93	.426	34
	Rockford Peaches	62	100	.383	41
	Humongous Melonheads	49	113	.302	54

Imaginary Baseball League Standings

Methods

Dataset Our team obtained data from the baseball resources *Fangraphs* and *Baseball Savant's Statcast*. Some of the following variables were used:

- DRS
- Age
- Position
- Innings/Game

Measuring Defensive Performance A fielder's **Defensive Runs Saved** (DRS) indicates how many runs a player saved or cost his team in the field compared to others at his position. This is the metric used to evaluate a player's ability to play a given position.

Some components of defensive runs saved:

- rSB: stolen base runs saved (catchers/pitchers)
- rGDP: double play runs saved (middle infielders)
- rARM: outfield arm runs saved (outfielders)

chart.jpg

Similarity Scores were introduced by Bill James in the mid-1980s. They are a method used to compare baseball players with each other, and are generally used to find the most similar historical players to a specific player. Currently, similarity scores are used for batting and pitching, and not for defense.

Modeling

In modeling the career trajectory of a player, two different models considered were a model that uses only the most similar players and a model that uses all defensive seasons at the same position.

Similarity Metric The similarity metric used is different than standard similarity score formulas because it doesn't attempt to weigh differences in defensive statistics. Instead, we calculate the distance between player-season vectors.

Mahalanobis Distance

The **Mahalanobis Distance** is a multi-dimensional measure of the distance between a point and a distribution, with consideration made to the correlation between the variables.

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}$$

[We might not need these bullet points]

- Scale-invariant
- Accounts for correlation between variables

Aging Models Traditional aging models in sports

Results

Comparing additional models produced the following results on accuracy:

Method	Acc.
Bagging	71.7%
Random Forest	73.0%
Logistic Regression	71.1%
Linear Disc. Analysis	72.0%
Support Vector Machine	74.8%

Additionally, the Random Forest model allowed us to find that the most important stats for predicting player positions were Putouts, Assists, the *Def* statistic from *Fangraphs*, and Age.

Conclusion

After Modeling the Data using the machine learning process, we found a model with a 73% classification rate. We learned how to work with variables, how to clean data, and identify trends.

Technologies Used/Skills Learned

- SQL
- R/Rstudio
- Excel
- Data Wrangling
- Data Visualization
- Mahalanobis Distance
- Sabermetrics
- Age Curves
- LaTeX

Acknowledgements

- University of Pittsburgh Department of Mathematics
- Dr. Jeffrey Wheeler
- Dr. Andrew Walsh of Health Monitoring Systems
- Joseph "Nico" Gabriel of Pitt School of Pharmacy
- Dr. Sam Ventura of the Pittsburgh Penguins

