The final recognition for professional baseball players deemed to have had outstanding careers is enshrinement in the Baseball Hall of Fame (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.

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 Database. 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).

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 Database. 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).

The Statistical Analysis System (SAS)® (Version 9.3, copyright 2010) was employed to create partial and final career totals for all variables. The Statistical Analysis System (SAS)® (Version 9.3, copyright 2010) was employed to create partial and final career totals for all variables. Logistic regression results found that hits, batting average, runs, batted in, and position played are important predictors of induction into the Hall of Fame.

For current players career end statistics were projected from partial career statistics. Logistic regression analysis identified significant variables in estimating the likelihood of enshrinement for selected current players. Logistic regression analysis correctly “fit” 95% of eligible retired players.

For pitchers important variables include strike outs, saves, wins, and ERA. The logistic regression analysis correctly “fit” 95% of eligible retired players.

For hitters important variables include hits, batting average, RBI, and position.

The logistic regression analysis correctly “fit” 95% of eligible retired players.

The logistic regression analysis correctly “fit” 95% of eligible retired players.

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.

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.

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

References