

Sabermetrics Project Report

Maxwell Lambek

May 7, 2018

Day of the week efficiency for teams

There are many ways to generate predictions for how well teams will perform against each other, and there are many statistics that can go into predicting a match outcome. Just comparing team win-rates gives a very general approximation that isn't consistently accurate. To make a better prediction one can compare win-rates and players performance at the park, but now more data is needed, and more calculations are needed. Adding more and more statistics will give a better prediction until overfitting starts to happen, and when that happens the high variance statistics should be removed or replaced, and this is where my statistic comes in. The statistic I made is a very simple with low variance between years for each team, and yet it can greatly help for predicting what teams are likely to beat other teams.

My statistic comes from the fact that we are creatures of habit, and tend to develop preferences and become comfortable in certain situations. A very simple preference that can develop on a team is preferring to play on certain days of the week. There are many more preferences that could be considered such as the time of day a team plays or weather conditions but, day of the week is a very simple model, and as a result has lower variance between years. The low variance can be seen without even calculating the sample variance because if a team had a significant amount of variance for a given day it would converge to the league average with a slight scaling from the team's general winrate.

My statistic isn't relevant for every team because some are just following

the league average or are statistically the same as other teams. To show if the desired team has a different mean for each day I used a T-test against all other teams and averaged the values. When I run my code for Boston I find that Boston is quite different when away from the league averages, and the T-test says that they are also quite different from each team (when playing away) in the league. So, my statistic would be a good addition to calculating win expectancy for Boston vs other teams. However, when looking at the Colorado Rockies the home-win rate by day is very unlikely to be of any use since it is very close to the league averages, and its T-test is nearly 0 with p value of 0.975 which isn't that large of a p value.

The necessary data to calculate my statistic can be obtained from Retrosheet. For each game Retrosheet has day of the week, score, and time of day, attendance, etc. which covers all the fields needed for day of the week and other preference and comfort based statistics.