

Modeling Playoff Success



Can regular season statistics be used to predict a playoff bracket?

Goals:

To figure out which regular season statistics, if any, can be used to accurately predict playoff success.

If such numbers do exist, build a model to predict series and test it against other data.

Data Organization and Collection

1. Find all of the Championship Series and World Series competitors from 1995 to 2015
2. Organize the regular season stats for those teams
3. Separate into a modeling group and a testing group
4. Stats gathered:
 - a. **Pitching:** ERA, FIP, WHIP, BB.9, HR.9, Pitching WAR, SV%, K.9, KBB
 - b. **Batting/Fielding:** Steal%, wRC, AVG, OBP, SLG, OPS, ISO, HR, Fld, R, WAR
 - c. **General:** Win%, Pythagorean Win %

Testing

For each statistic, the code:

1. Looks to see which team in the matchup has a better score for that variable; that team is the “favorite”.
2. Checks to see if the “favorite” won
3. Records the percentage of “favorite wins” for each category

Results

	Win.Percent.by.var ▼
SV..	0.579
wRC	0.561
OPS	0.561
R	0.561
wRC.1	0.561
Pythag	0.544
WAR.1	0.544
K.9	0.509
HR	0.491
WHIP	0.491
HR.9	0.491
SLG	0.474
Win..	0.456

BB.9	0.456
Steal..	0.439
Steal...1	0.439
K.BB	0.439
WAR	0.439
ERA	0.439
FIP	0.439
ISO	0.421
Fld	0.421
OBP	0.404
AVG	0.316

Building a Projection Model

1. The untested sample is run through; each team is:
 - a. Assigned points for being superior in each category, where the point value is the probability from earlier
 - b. The proportion of points assigned to each team is the percentage chance of its victory

Testing the Projection Accuracy

A **Brier Score** measures the accuracy of a percentage-based prediction

The worst possible score is a 1, and the best is a 0.

Our Brier Score is a 0.263

Brier Scores

2015	Royals	AL	0.17471196	0.26300925
2015	Mets	NL	0.273999	
2015	Blue Jays	AL	0.15098595	
2015	Cubs	NL	0.21561953	
2014	Giants	NL	0.38428109	
2014	Royals	AL	0.26437699	
2014	Cardinals	NL	0.37904647	
2014	Orioles	AL	0.57340363	
2013	Red Sox	AL	0.74640753	
2013	Cardinals	NL	0.33794944	
2013	Tigers	AL	0.30545286	
2013	Dodgers	NL	0.03692789	
2012	Giants	NL	0.229441	
2012	Tigers	AL	0.253481	
2012	Cardinals	NL	0.349281	
2012	Yankees	AL	0.395641	
2011	Cardinals	NL	0.114244	
2011	Rangers	AL	0.303824	
2011	Brewers	NL	0.009604	
2011	Tigers	AL	0.161604	
2010	Giants	NL	0.303601	
2010	Rangers	AL	0.250049	
2010	Phillies	NL	0.063504	
2010	Yankees	AL	0.351649	
2009	Yankees	AL	0.060809	
2009	Phillies	NL	0.1589585	
2009	Angels	AL	0.041616	
2009	Dodgers	NL	0.164836	

2008	Phillies	NL	0.63308575	
2008	Rays	AL	0.39368638	
2008	Dodgers	NL	0.7520825	
2008	Red Sox	AL	0.17043208	
2007	Red Sox	AL	0.19229402	
2007	Rockies	NL	0.25003541	
2007	Indians	AL	0.20054526	
2007	Diamondback	NL	0.32443858	
2006	Cardinals	NL	0.22915544	
2006	Tigers	AL	0.17158779	
2006	Mets	NL	0.13995261	
2006	Athletics	AL	0.10620925	
2005	White Sox	AL	0.52392291	
2005	Astros	NL	0.51690731	
2005	Angels	AL	0.22528038	
2005	Cardinals	NL	0.25932915	
2004	Red Sox	AL	0.27601934	
2004	Cardinals	NL	0.23250573	
2004	Yankees	AL	0.1439823	
2004	Astros	NL	0.09175023	
2003	Marlins	NL	0.11406365	
2003	Yankees	AL	0.26750212	
2003	Cubs	NL	0.04263957	
2003	Red Sox	AL	0.36376845	

The Coin Flip Test

A randomly flipped coin has a Brier Score of 0.25;

So, a **coin flip** is a **more accurate** predictor of playoff success

Why?

The probability of victory chart works only for the past.

A regression on our “most influential” factors have miniscule regressions and large P-Values.

So, the correlations between certain statistics and victory is not at all predictive.

Takeaways

Playoffs are very hard to predict based on season stats.

Because of all the randomness involved, a team should strive to just get into the playoffs in the first place.

Even the worst playoff team has a chance of winning the World Series.