

World Series Champions Simulation

Paul Forst, Ian O'Connor, Nathan Thomas

March 14, 2018

The Rundown

This simulation collected statistics for the 16 Major League Baseball World Series champions from 2001 through 2016, created a game simulation and paired the teams up in games to see which team is the best. The batting and pitching statistics are included for all teams from their respective championship seasons.

To determine a true champion, the 16 teams will play a simulated season, playing all other teams twice, once at home and once away, in a round-robin format season. The simulation will run for 100 “seasons” and output a list of every season winner and their record in that particular season.

Probabilities for the simulation were based on modified “on-base” percentages which calculated the probability that any given batter would obtain one base, two bases, three bases, or four bases in an at bat or get out. The top nine batters from each team were used to create a lineup. These nine batters were selected by choosing the top 9 with the highest at bats through their respective regular season. Pitching was factored into the simulation by calculating a similar “on-base” percentage, whereby producing an “apples-to-apples” environment for the simulation. The on-base percentages were then adjusted by the average on-base percentage for pitchers on all 16 teams rosters. This adjusted score was then weighted by the number of batters the pitcher faced throughout the season. The scores were then aggregated creating a new base score for the entire pitching staff. The new base scores were averaged and then the average was reduced from the base score creating a factor. Negative factors meant that a pitching staff was very good and would have a negative effect on the hitters probabilities, while positive factors meant that a pitching staff was not as good and hitters would fare much better against the pitching staff.

During the game simulation, the factor for the opposing pitching staff was multiplied against the hitting team’s on-base probabilities. The simulation for each game is then run.

```
## NULL
```

```
## # A tibble: 10 x 8
```

##	Team	Record	HomeRec	AwayRec	WinPct	RunsFor	RunsAgainst	Last10
##	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<chr>
##	1	NYA_2009	25-5	14-1	11-4	0.833	391	241 9-1
##	2	BOS_2004	19-11	9-6	10-5	0.633	374	256 5-5
##	3	BOS_2004	23-7	11-4	12-3	0.767	470	314 6-4
##	4	SLN_2006	21-9	11-4	10-5	0.700	353	283 10-0
##	5	BOS_2004	23-7	11-4	12-3	0.767	448	222 7-3
##	6	BOS_2007	19-11	5-10	14-1	0.633	306	246 3-7
##	7	SLN_2011	25-5	14-1	11-4	0.833	344	213 9-1
##	8	BOS_2004	21-9	12-3	9-6	0.700	395	276 8-2
##	9	BOS_2004	22-8	10-5	12-3	0.733	436	252 5-5
##	10	NYA_2009	24-6	12-3	12-3	0.800	394	227 8-2

Final Results

Once the simulation is finished running and we have each season’s winner, we can aggregate these and output the overall champion, or the team that won the most simulations.

```
## [1] "Your champion is BOS_2004"
```