# Notes for Super Rookies vs. '27 Yankees sim

The data are saved as a .csv file. Convert to preferred data file type for statistical analysis (.dta for Stata).

There are two simulation scripts: Rookies_Yankees and Generic. The first is specific to the Rookies vs. '27 Yankees series featured in the blog post. The second enables the underlying simulation to be run for any two teams once their respective *WAR per game* values (team WAR/games played) are input.

Both scripts are written as Stata .do files. However, the annotations explain the operations with sufficient precision to enable the routines to be coded in alternative languages.

Two key variables are simulated. The first is the relative strength of the teams and the resulting 1-game win probability for the Super Rookies (or Team 1 in the generic script). Random values are drawn for each team from the probability distribution associated with the regression-model estimates of their respective winning-percentages against average opponents. Based on those, a 1-game win probability is determined using Bayes's Theorem. The second simulated value is the outcome of a game between the teams. A random number between 0 and 1 is generated; whether that number falls above or below the 1-game win probability previously generated determines the outcome of that game. As described in the blog post, this process is repeated to determine the outcome of each game of 2,500 best-of-7 series.

The generated output indicates the range of simulated 1-game win probabilities for the Rookies (or Team 1); the number of series won by each team; and the proportion of series outcomes that resulted in victories for one team or the other in 4, 5, 6, or 7 games.