

wOBA user notes

The data files are either .csv or .xlsx, and should be converted to the form suitable for statistical analysis in the statistics package of your choice (.dta for Stata).

The analysis script is a .do file. But it is annotated sufficiently to enable translation to other programming languages.

Single-season regression models use season-specific computations of runs per game and of the batting-production measures.

The multi-season regression analyses use season-standardized forms of these metrics and of the outcome variable runs-per-game (“rpg”). Standardization by season enables multi-season measurement of how the relevant metrics influence *season-specific* variance in runs scored (as opposed to across-season variance, which is not a quantity of much relevance). It also is an alternative to multi-level modeling or time-period regressions for removing extraneous variability associated with shifting game conditions over time. See the discussion in Schell, M. J., *Baseball's All-Time Best Hitters: How Statistics Can Level the Playing Field* (Princeton University Press 2013); Schell, M. J., *Baseball's All-Time Best Sluggers: Adjusted Batting Performance from Strikeouts to Home Runs* (Princeton University Press 2016).

The script in its present form retains code that was used to construct a multivariate-regression version of wOBA. Don't be taken in by the R^2 ; an examination of the coefficients associated with the models it generates across seasons will reveal why this approach was abandoned.