

Notes for Standardized Batting Averages

Note that the program uses only “modern era”—i.e., post-1899—AL/NL performance data.

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

The data script is a Stata .do script. The annotations explain the operations with sufficient precision to enable the routines to be coded in alternative languages.

The exported excel file will contain many variables that were generated to produce the standardized batting averages. They are retained because they may be of interest and of utility for refining or extending the program. But they will constitute clutter if the goal is to produce a clean output of the results. To do so, you will need to either delete or hide various columns. You will also likely want to move to the far left the Columns AJ and AK, which supply the full names (as opposed to the Lahman Database codes) of the batters. Also, you will want to format cells that include numerical outputs of interest (e.g., player batting averages, BA+ scores, and batting average z-scores) so that they display the desired number of decimal points.