

Statistics 4590: Baseball Salaries and Performance

We always seem to read discussion about the salaries of professional athletes in the paper. One often wonders if salaries are related to how a player performs. The focus of this project is on salaries of baseball players.

The data set is on the course website, and you can use the `read.table` command to read the data into R.

Your goal is to develop an appropriate model to predict a player's salary based on his performance, and measures of his ability to switch teams (free agency) or adjust his salary (arbitration). In your analysis, you should try to determine as uncomplicated a model as possible, but one that still has reasonable predictive skill. You also want to see if your model assumptions are satisfied (is any data transformation needed?), and whether any outliers or influential observations are having a dramatic affect on your estimates.

The salary data come from the 1992 season, while the player statistics come from the 1991 season. The data file has the following form:

1. Player's name
2. Salary (in thousands of dollars)
3. Batting average
4. On-base percentage (OBP)
5. Number of runs
6. Number of hits
7. Number of doubles (Dbls)
8. Number of triples
9. Number of home runs (HR)
10. Number of runs batted in (RBI)
11. Number of walks
12. Number of strike-outs (SO)
13. Number of stolen bases (SB)
14. Number of errors (Err)
15. Indicator of "free agency eligibility" (1 if eligible) (FAEL)
16. Indicator of "free agent in 1991/2" (1 if free agent) (FA9192)
17. Indicator of "arbitration eligibility" (1 if eligible) (AREL)
18. Indicator of "arbitration in 1991/2" (1 for arbitration) (AR9192)