Robust Sabermetrics

L. McCann[1], R. Welsch[1], and C. Liberty[1]

[1] Massachusetts Institute of Technology, Sloan School of Management, 77 Massachusetts Avenue, Room E53-383, Cambridge, MA 02139

Keywords: robust regression, selection, diagnostics, sports statistics.

Abstract

A recent article by Lock (2004) in *STATS* discusses the role extreme values and unusual data might play in the relationship between how many wins a major league baseball team has in a year and the team's player payroll. The article concludes that extreme values do not seem to make much difference in many cases when using methods like linear regression and correlation, rank correlation, difference in means, and the Mann-Whitney test. Still euphoric from the recent Boston World Series triumph over the Yankees (which has by far the largest payroll), we decided to take a closer look.

Sabermetrics (from SABR, the Society for American Baseball Research) is the term used by Michael Lewis (2004) in his book *Moneyball* to describe the search for new baseball knowledge using statistics. The Red Sox are reportedly avid users of these methods.

In this paper, we apply modern robust and diagnostic methods to the types of data discussed in *Moneyball* and the *STATS* article. We first begin with simple diagnostics and diagnostic traces, move on to classical robust methods and then use the best robust methods we can find in widely available commercial software and in free public program libraries. We conclude with some new approaches based on least angle regression (Efron et al. 2004), identity matrix augmentation for simultaneous variable selection and robust estimation, sampling, and cross validation along with a discussion of issues that arise when the number of variables is larger than the number of observations.

References

B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least Angle Regression. *The Annals of Statistics*, 32, 2, 407–499.

M. Lewis (2004). *Moneyball: The Art of Winning an Unfair Game.* W.W. Norton & Company.

R. Lock (2004). Can Major League Baseball Teams Buy Success? *STATS: The Magazine for Students of Statistics*, 39, 22–23.