Data Analysis Exercises for Chapter 22: *Applied Regression Analysis, Generalized Linear Models, and Related Methods, Third Edition* (Sage, 2016)

John Fox

Last modified: 2015-01-29

Exercise D22.1 The data file `BaseballPitchers.txt` contains salary and performance data for major-league baseball pitchers at the start of the 1987 season. The data are analogous to those for baseball hitters used as an example in the chapter. Employing one or more of the methods of model selection described in the text, develop a regression model to predict pitchers' salaries. Be sure to explore the data and think about variables to use as predictors before specifying candidate models. How successful is the model in predicting salaries? Does the model make substantive sense?

Exercise D22.2 Repeat Exercise D22.1, but this time divide the data randomly into two subsamples, applying one or more methods of model selection to the first (training) subsample. Then evaluate the selected models on the second (validation) subsample. What do you find? If you employed more than one method of model selection, how do they compare?

Exercise D22.3 Apply Bayesian model averaging to the baseball-pitchers salary data from Exercise D22.1. Compare the results to those obtained in Exercise D22.1