Data Analysis Exercises for Chapter 12: *Applied Regression Analysis, Generalized Linear Models, and Related Methods, Third Edition* (Sage, 2016)

John Fox

Last modified: 2015-01-29

Exercise D12.1 In Exercises D5.5/D6.5, D.7.1, and D8.2 various linear models were fit to data. Use the methods of this chapter to check for non-normality, non-constant error variance, and (as appropriate) nonlinearity in each of these analyses. In each case, attempt to correct any problems that are detected. Because many different methods are discussed in this chapter, you might find the following strategy useful: Use relatively simple diagnostics to check for problems and more sophisticated methods to follow up. To check for non-normality, construct a quantile-comparison plot and a kernel density estimate or histogram of the studentized residuals; to check for non-constant error variance, plot studentized residuals against fitted values; to check for nonlinearity, examine component+residual plots.

Exercise D12.2 Canadian occupational prestige regression:

(a) In Section 5.2, occupational prestige is regressed on income, education, and percent women for 102 Canadian occupations. Repeat Exercise D12.1 for this regression.

(b) Consider the following alternative analysis of the Canadian occupational prestige data: Regress prestige on income, education, percent women, and on dummy regressors for type of occupation (professional and managerial, white collar, blue collar); include interactions between type of occupation and each of income, education, and percent women. Why is it that the interaction between income and type of occupation can induce a nonlinear relationship between prestige and income when the interaction is ignored? [*Hint*: Construct a scatterplot of prestige vs. income, labeling the points in the plot by occupational type, and plotting the separate regression line for each occupational type; see Section 7.3, where a similar analysis is performed but omitting percent women.]