4.7 Model Selection

Model Selection

The `BigMac2003` data set can be found in the package `alr4` in R (source: Cook and Weisberg, “Applied Regression Including Computing and Graphics,” Wiley, 1999)

The Big Mac hamburger is a simple commodity that can be used to study the inefficiency in currency exchange, see an article in the Economist.

(a) Confirm that a log-transformation is appropriate for all variables which are measured in units of currency (minutes of labor).

(b) Assume the log-price of a BigMac as the response and carry out a best-subset linear regression analysis.

Compute the AIC, BIC, five- and tenfold cross-validation of prediction error for the best model and the full model. Discuss the results. (Hint: you may use the package "bestglm".)

(c) Compare the diagnostic plots for the chosen model and the full model, e.g. by overlaying each plot. Which cities are most influential for the fits? Are there any outliers?

(d) Assuming that it is unknown, give a confidence interval and a prediction interval for the price of a BigMac in Sydney. Which model do you suggest using for the prediction?