(1) Obtain $PRESS_p$ for the three best models from homework # 5: the best 4-, 5- and 6-variable models for the even-numbered observations in the CDI data set). What do these $PRESS_p$ values indicate? How do they compare with the results of the other selection criteria used in Homework # 5?

(2) 9.28 Refer to the CDI data set. Use the data set of the odd-numbered CDIs (observations with odd ‘IDNum’) as a validation data set to validate the best 4-variable model identified in Homework # 5: the model with % 18-34, # beds, % poverty, and % unemployment as the explanatory variables.

a. Fit the regression model to the validation data set. Compare the estimated regression coefficients and their estimated standard deviations with those obtained for the model-building data set. Also compare the error mean squares and coefficients of multiple determination. Does the model fitted to the validation data set yield similar estimates as the model fitted to the model-building data set?

b. Calculate the mean squared prediction error in (equation 9.20 [5th ed.] or 10.60 [4th ed.]) and compare it to MSE obtained from the model-building data set. Is there evidence of a substantial bias problem in MSE here?

c. Fit the selected regression model to the combined model-building and validation data sets. Are the estimated regression coefficients and their estimated standard deviations appreciably different from those for the model fitted to the model-building data set? Should you expect any differences in the estimates? Explain.