Homework # 5

due Wednesday, 19 February

in discussion

(1) 9.26 with modifications:
Refer to the CDI data set. A public safety official wishes to predict the rate of serious crimes in a CDI (Y, total number of serious crimes per 100,000 population). The pool of potential predictor variables includes all other variables in the dataset except total population, total serious crimes, county, state, and region. It is believed that a model with predictor variables in first-order terms with no interactions terms will be appropriate. Consider the even-numbered cases to constitute the model-building data set to be used for the following analyses.

a. Prepare separate histograms for each of the predictor variables. Are there any noteworthy features in these plots? Comment.

b. Obtain the scatter plot matrix. Also obtain the correlation matrix of the X variables. Is there evidence of strong linear pairwise associations among the predictor variables here?

c. [modified entirely]
Use \( C_p \) and \( R^2_{a,p} \) in an all-possible-regression procedure.
Calculate or obtain \( AIC_p \) and \( SBC_p \) for the best (largest \( R^2 \)) 4-, 5- and 6-variable models.
Compare the results of these four criteria (\( C_p \), \( R^2_{a,p} \), \( AIC_p \) and \( SBC_p \)).
Select 3 to 5 models you think deserve further evaluation.

DATA NOTES:
The response variable you are to use is the number of serious crimes per 100,000 people: total serious crimes / (total population / 100000).
You are to use only the even-numbered cases.