Statistics 151a Spring 2015 (Linear Modelling -Theory and Applications) Homework Four

Due on 13 April, 2015

02 April, 2015

1. Consider the bodyfat dataset and consider fitting a linear model for the response variable BODYFAT in terms of the explanatory variables AGE, WEIGHT, HEIGHT, ADIPOSITY, NECK, CHEST, ABDOMEN, HIP, THIGH, KNEE, ANKLE, BICEPS, FOREARM and WRIST.

   (a) Using each of the following methods, perform variable selection to select a subset of the explanatory variables for modeling the response:

      i. Backward elimination using the individual p-values.

      ii. Forward Selection using p-values.

      iii. Adjusted R^2.

      iv. AIC

      v. BIC

      vi. Mallow's C_p.

   (b) Let $M_1, \ldots, M_6$ denote the six models selected by each of the six variable selection methods of the previous part. Select one of these models by cross-validation.

   (c) Let M be the model selected in the previous part. Fit this model to the data. Perform regression diagnostics. Comment on the validity of the assumptions of the linear model. Identify influential observations and outliers. Delete them if necessary and re-fit the model.