STA244 **4/7/2003**

Homework 7

Due 4/14/2001

1. The matrix $X'_{(i)}X_{(i)}$ can be written as $X'_{(i)}X_{(i)} = X'X - x_i x'_i$ where x'_i is the ith row of X and $X_{(i)}$ is the matrix X with the ith row removed. Use this to show

$$(X'_{(i)}X_{(i)})^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}x_i x'_i (X'X)^{-1}}{1 - h_{ii}} \tag{1}$$

2. The quantity $Y_i - x'_i\hat{\beta}_{(i)}$ is the residual for the ith case when β is estimated without the ith case. Show that

$$Y_i - x'_i\hat{\beta}_{(i)} = \frac{\hat{e}_i}{1 - h_{ii}} \tag{2}$$

3. Show that Cook's distance (page 365) can be written as

$$D_i = \frac{1}{p} r_i^2 \frac{h_{ii}}{1 - h_{ii}} \tag{3}$$

   where $r_i = \hat{e}_i / \sqrt{\hat{\sigma}^2(1 - h_{ii})}$ is the internally Studentized residual.

4. The data set in
   `http://www.stat.duke.edu/courses/Spring03/sta244/Data/rat.dat`
   contains results from an experiment conducted to investigate the amount of drug retained in the liver of a rat. Nineteen rats were randomly selected, weighed, placed under light anesthesia and given an oral dose of the drug. The dose an animal received was determined as approximately 40 mg of the drug per kilogram of body weight, since the liver weight is known to be strongly related to body weight and it was felt that large livers would absorb more of a given drug than smaller livers. After a fixed period of time, each rat was sacrificed, and the liver weighed, and the percent dose (Y) in the liver determined.

   The researchers' hypothesis was that for the method of determining the dose given to each rat, there there is no relationship between the percentage of the dose in the liver (Y) and the body weight in grams (Body), liver weight in grams (Liver), and relative dose (Dose). Fit statistical models to the data to examine their hypothesis. Write a 1 page (typed) report describing your analysis and conclusions, with 1-2 pages max of supporting figures and tables. Any figures/tables should be labeled and referenced within the body of the report.

5. Extra Credit: Write a function to implement the Bayesian outlier Diagnostics in R or another language. Verify that it works correctly with the stackloss data.

   `http://www.stat.duke.edu/courses/Spring03/sta244/Data/stackloss.txt`