STA244 **4/9/2002**

Homework 7

Due 4/16/2001

1. Problem 15.7 in CW.

   To obtain case diagnostics in S-Plus, fit a model using the QR option, i.e.
   `mylm.obj <- lm(Y ~ X1 + X2, data=mydataframe, qr=T)`

   To obtain the case diagnostics, use the function `ls.diag()` in the command line:
   `mylm.diag <- ls.diag(lm.obj)`
   This will create an object `mylm.diag` that contains

   - leverage `mylm.diag$hat`,
   - Cook's distance `mylm.diag$cooks`
   - Externally Studentized residuals `mylm.diag$stud.res`

   Download Bayesian outlier functions from the web page.

2. 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 prove

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

3. The quantity $Y_i - x'_i\hat{\beta}_{(i)}$ is the residual for the ith case when β is estimated without the ith case. Use the above result to show that

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

4. Using (1), verify that Cook's distance can be written as

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

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

5. Use (1) to prove the relationship between r_i and t_i in equation 15.14 in CW. Show that the two approaches for testing outliers outlined in the text (using the t_i in Equation 15.11 and testing $\delta = 0$ in the model in 15.12) are in fact equivalent. *Hint:* you qill need to prove that $x'_i(X'_{(i)}X_{(i)})^{-1}x_i = h_{ii}/(1 - h_{ii})$

6. Prove that

$$\hat{\sigma}^2_{(i)} = \hat{\sigma}^2 \frac{n - p - r_i^2}{n - p - 1}$$