STAT 3A03 Applied Regression With SAS

Assignment 3 **Revised Due Date: 5pm on Friday November 3, 2017**

Dropboxes for assignment submission are outside HH-105. Your assignment **MUST** be deposited in the appropriate dropbox for your lab section.

N.B. Late assignments will not be accepted

Q. 1 The Scottish Hill Races dataset describes the fastest times (in seconds) in 35 races in the Highlands of Scotland as well as the length of the races (in miles) and the vertical climb (in feet). The dataset is available on the website in the file `ScottishHillRace.txt`.

a) Fit a linear model relating the record time to distance and climb.

b) Use the appropriate diagnostics to check if any of the model assumptions are violated and comment on your results.

c) Use plots of Cook's Distance, the Hadi measure and the potential-residual plot to determine if there are any points which are suspect. Identify the races for any such points and explain what the issue is with each point.

d) Remove each of the points identified in part (c) in turn and refit the model. For each point explain the effect of removing that point on the estimates of the model parameters (β_0, β_1 and σ^2).

Q. 2 Textbook 4.12 (The data are in `Table4_8.txt`)

Q. 3 Textbook 4.13 (The data are in `Table4_8.txt`)

Q. 4 **a)** Give a careful definition of jackknifing and explain why it is an important concept in linear regression.

b) Show the following results relating quantities obtained from a complete data set with those when observation i is omitted in a simple linear regression context.

(i) $\overline{x} = \overline{x}_{(i)} + \dfrac{x_i - \overline{x}_{(i)}}{n}$.

(ii) $S_{xx} = S_{xx(i)} + \dfrac{n-1}{n}\left(x_i - \overline{x}_{(i)}\right)^2$.

(iii) $S_{xy} = S_{xy(i)} + \dfrac{n-1}{n}\left(x_i - \overline{x}_{(i)}\right)\left(y_i - \overline{y}_{(i)}\right)$

c) Use the results from part (b) to give formulae for the least squares estimates of the intercept and slope in the model without observation i using only statistics obtained using all observations ($\overline{x}$, $\overline{y}$, S_{xx}, S_{xy}) and the values of x_i and y_i.