Queen Mary University of London

School of Mathematical Sciences

You should hand in answers to **Questions 2 and 3 only** by **Thursday, 31 January, 3.00 pm.** *Please put your course-work into the* ***yellow box*** *on the* ***Basement*** *level of Maths Building. The work should be stapled if necessary.*

Question 1. Dr Joseph Hooker collected a set of 31 measurements on the boiling point of water (BP in degrees Fahrenheit) and the atmospheric pressure (AP in inches of mercury) at various locations in the Himalaya Mountains. The data are given on the course web-page in the file BoilingPoint.txt

(a) Draw a scatter plot of the data. Do the data suggest a relationship between temperature of boiling water and pressure in the Himalayas?

(b) Obtain the simple linear regression model fit and the standardized residual plots.

(c) Do the residuals contradict the linearity of the model and the assumption of normality of the random errors? Explain. (*Note that the linearity may be examined by checking the residuals versus fitted values as well as versus values of the explanatory variable.*)

Question 2. For the data in Question 1 transform the explanatory variable to a new one to get $Z = 100\ln(AP)$ and then

2.1 Repeat items (a) - (c) from Question 1 using Z as the explanatory variable. Did the residuals improve? Is there a contradiction against the constant variance assumption?

2.2 Interpret the result of the F-test given in the ANOVA table.

2.3 Suppose that the temperature had been measured in $^\circ C$ instead of $^\circ F$. Derive the relationship between the Least Squares estimator of β_1 for the two scales. Do not use MINITAB for this question.

Question 3. For a simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where $\varepsilon_i \underset{iid}{\sim} N(0, \sigma^2)$, show that the variance of the crude residuals $e_i = Y_i - \widehat{Y}_i$ is given by

$$\mathrm{var}(e_i) = \sigma^2 \left[1 - \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)\right].$$

(*Note that* $\sum_{j=1}^n c_j = 0$, $\sum_{j=1}^n c_j x_j = 1$ and $\sum_{j=1}^n c_j^2 = \frac{1}{S_{xx}}$, *where* $c_j = \dfrac{x_j - \overline{x}}{S_{xx}}$.)