QUEEN MARY UNIVERSITY OF LONDON

MTH5120 Statistical Modelling I

Exercise Sheet 11

1. We use the Bridge.txt dataset available on QMPlus, where information from 45 bridge projects are compiled. The response and predictor variables are as follows:

   - Y: Time is the design time in person-days;
   - X_1: DArea is the deck area of bridge (000 sq ft);
   - X_2: CCost is the construction cost ($000);
   - X_3: Dwgs is the number of structural drawings;
   - X_4: Length is the length of bridge (ft);
   - X_5: Spans is the number of spans.

   Take the logarithm transformation of all the variables.

   (a) By using the AIC find the best two models. For the two best models, find the leverage values and the Cook's distance values and comment on them.

2. When fitting the model

$$E[Y_i] = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

   to a set of $n = 5$ observations, the following results were obtained using the general linear model notation:

$$\left(\boldsymbol{X}^t \boldsymbol{X}\right)^{-1} = \begin{pmatrix} 209.32 & -3.82 & -0.71 \\ -3.82 & 0.069 & 0.013 \\ -0.71 & 0.013 & 0.002 \end{pmatrix}$$

   with variables:

Y	X_1	X_2
92.5	50.9	20.8
94.9	54.1	16.9
89.3	47.3	25.2
94.1	45.1	49.7
98.9	37.6	95.2

   (a) Have any of the observations high leverage value?

(b) Let us consider the problem

$$E[Y_i] = \beta_0 + \beta_1 x_{1,i}$$

with Y and X_1 defined as above with the exception of $x_{1,5}$, which changes from 37.6 to 20. Thus the matrix $\left(\boldsymbol{X}^t\boldsymbol{X}\right)^{-1}$ becomes

$$\left(\boldsymbol{X}^t\boldsymbol{X}\right)^{-1} = \begin{pmatrix} 2.77 & -0.06 \\ -0.06 & 0.001 \end{pmatrix}$$

have any of the observations high leverage value?