STAT 3008
Homework 4
Due date: 5pm, Dec 5 (Friday). Assignment Box at LSB 1/F.

1. Consider the dataset
   $y = (4.562, 12.280, 21.018, 35.643, 54.084, 76.642, 192.780)$,
   $x1 = (1, 2, 3, 4, 5, 6, 7)$,
   $x2 = (4, 3, 5, 2, 6, 1, 7)$.

   i) Draw a scatterplot matrix to represent the dataset.

   ii) Use Box-Cox transformation to select a model (transform response only). Select among $\lambda = (-1, -1/2, 0, 1/3, 1/2, 1, 2)$.

   iii) Use modified power transformation to select a model (transform predictors only). (There are a lot of possibilities, just target at one transformation that you feel reasonable.)

   iv) Compute the RSS for each of the best fit model in ii) and iii). Which method gives a smaller RSS?

2. Consider the weighted least square regression

$$Y = X\beta + e, \qquad e \sim N(0, \sigma^2 W^{-1}), \tag{1}$$

   where X is a $n \times (p+1)$ matrix.

   i) Find the hat matrix H for weighted least square regression, where H satisfies $\hat{Y} = HY$.

   ii) Is H symmetric?

   iii) Is $HH = H$?

   iv) Is $HX = X$?

   v) Is $X'H = X'$?

   vi) Is $tr(H) = p + 1$?

3. For the regression model $Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + e$, $e \sim N(0, \sigma^2 I)$, find

   i) $E(\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2)$.

   ii) $E(\sum_{i=1}^n (Y_i - \hat{Y}_i)^2)$.

   iii) $E(\sum_{i=1}^n (Y_i - \bar{Y})^2)$.

4. Consider the following dataset.

	Y	X_1	X_2	X_3
1	2	1	1004	5.2
2	1	200	806	6.1
3	3	-50	1058	4.9
4	4	909	941	6.5
5	1	506	100	2.4
6	8	1200	505	8.1

   i) Find the variance inflation factor for each of the predictors.

   ii) Use the Forward Selection algorithm with AIC criterion to select the best model.

   iii) Use the Backward Elimination algorithm with BIC criterion to select the best model.