HW 4
Due: Thursday, October 17, ’02

1. An experiment involved a quantitative analysis of factors found in high-density lipoprotein (HDL) in a sample of human blood serum. Three variables though to be predictive or associated with HDL measurement (Y) were the Total cholesterol (X_1) and the total triglyceride (X_2) concentrations in the sample, plus the presence or absence of a certain sticky component called sinking pre-beta, or SPB (X_3), which was coded as 0 if absent and 1 if present. The data obtained are stored in http://www.math.umbc.edu/~anindya/Stat601/F2002/hw4.q1.dat.

(a) a. Test whether X_1, X_2 and X_3 alone significantly helps in predicting Y.
(b) Test whether X_1, X_2 and X_3 taken together significantly helps to predict Y.
(c) Consider the model

\[ Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_{13}X_1X_3 + \beta_{23}X_2X_3 + \epsilon \]

i. Is this a LINEAR MODEL?
ii. Test the hypothesis \( H_0 : \beta_{13} = \beta_{23} = 0 \) vs \( H_A : \) at least one is nonzero.
iii. Write down in WORDS what it means to reject this test.

(d) Test whether X_3 is associated with Y, after the combined contribution of X_1 and X_2 are taken into account. What does your result, together with you answer from part (3), tell you about the relationship of Y with X_1 and X_2 when SPB is present, as compared with when it is absent.

(e) Load the data into X-GOBI and look at all the projections. Comment on your findings. [OPTIONAL]

2. The dataset http://www.math.umbc.edu/~anindya/Stat601/F2002/hw4.q2.dat contains 1000 values of \( y_i \). Suppose the population model is

\[ y(x) = f(x, \theta) + e(x), \quad x \in (0,1] \]

where \( f \) is a smooth function depending on a finite dimensional parameter \( \theta \). Suppose \( y_i = y(i/1000) \) for \( i = 1, 2, \ldots, 1000 \). Let us model \( y_i \) as

\[ y(x) = \beta_0 + \beta_{11}x + \beta_{12}x^2 + \beta_{13}x^3 + \beta_{21}(x - A_1)_+ + \beta_{22}(x - A_1)_+^2 + \beta_{23}(x - A_1)_+^3 + \beta_{31}(x - A_2)_+ + \beta_{32}(x - A_2)_+^2 + \beta_{33}(x - A_2)_+^3 + e(x) \]

where \( 0 < A_1 < A_2 < 1 \) and \( x_+ = 0 \) if \( x < 0 \) and equal to \( x \) if \( x \geq 0 \).

(a) Assuming \( A_1 = 0.3 \) and \( A_2 = 0.7 \) estimate the model using OLS.
(b) Using partial-t tests choose a sub model of (1) that you think adequately describes the data.
(c) Give some heuristics how you will choose the constants \( A_1 \) and \( A_2 \) if they were unknown.

3. Problems 3.4 and 3.5 from the text.