STAT 3008
Homework 3
Due date: Nov 18 (Tuesday). Assignment Box at LSB 1/F.

1. Let $Y = (21, 25, 19, 34, 36, 36, 24, 10)'$, $X1 = (3, 9, 4, 3, 7, 9, 4, 1)'$, $X2 = (3, 9, 4, 3, 7, 9, 4, 2)'$. Consider a multiple linear regression with an intercept term. At the point $(X1, X2) = (12, -2)$, estimate $E(Y|X1, X2)$ and find
   a) 90% confidence interval for the fitted value of $(X1, X2)$.
   b) 95% prediction interval for a new observation of $(X1, X2)$.
   c) 99% confidence band for the regression line at $(X1, X2)$.
   d) Find an inequality representing the 95% confidence ellipse for β.

2. For each case, construct a small data set consisting of one response and two predictor variables so that in the following two fitted equations,
$$\hat{Y} = \hat{\beta}_o + \hat{\beta}_1 X_1 \,, \quad \hat{Y} = \hat{\alpha}_o + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2 \,,$$
   i) the regression coefficient of X_1 are equal.
   ii) the regression coefficient of X_1 have opposite signs.

   (Hints: Try different covariance between X_1 and X_2.) For each case, give a practical example for Y, X_1 and X_2 that explains the behavior of fitted coefficients.

3. Suppose when the god created the world he designed that the kinetic energy y and the velocity x of an object follows the relation
$$y = \beta_o + \beta_1 x^2 + e \,, \quad e \sim N(0, \sigma_o^2) \,.$$
   Suppose you observed (x_i, y_i) for $i = 1 \ldots, n$, and you decided to fit the simple linear regression model
$$y = \alpha_o + \alpha_1 x + e \,, \quad e \sim N(0, \sigma^2) \,.$$
   for the data set.
   i) Using matrix notations (define the X and Y you use), find $\hat{\alpha} = (\hat{\alpha}_o, \hat{\alpha_1})$.
   ii) Using matrix notations, find the expected value $E(\hat{\alpha})$ and the variance $Var(\hat{\alpha})$.
   iii) Express $(E(\hat{\alpha}_o), E(\hat{\alpha_1}))$ in terms of $\sum_{i=1}^n x_i$, $\sum_{i=1}^n x_i^2$, $\sum_{i=1}^n x_i^3$ and (β_o, β_1).
   iv) If $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i = \mu_x$, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i^2 = \sigma_x^2$ and $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i^3 = \eta_x$, express $(\hat{\alpha}_o, \hat{\alpha}_1)$ in terms of (β_o, β_1), μ_x, σ_x^2 and η_x, when $n \to \infty$. Is $(\hat{\alpha}_o, \hat{\alpha}_1)$ a consistent estimator of (β_o, β_1)?

4. Consider the straight-line model $y = \beta_o + \beta_1 X + e$. Give the appropriate weight matrix W to use in a weighted least squares regression if the variance of the random error e_i, i.e., σ_i^2, is proportional to
   i)x_i^2, ii)$\sqrt{x_i}$, iii)x_i, iv) $1/n_i$, where n_i is the number of observations at level x_i.

5. Suppose you are investigating allegations of sex discrimination in the hiring practices of a particular firm. An equal-rights group claims that females are less likely to be hired than males with the same background, experience, and other qualifications. Data (see the table) collected on 12 former applicants will be used to fit the model $E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$, where
$$Y = \begin{cases} 1 & \text{hired} \\ 0 & \text{not} \end{cases} \quad X_1 = \text{Years of education}, \; X_2 = \text{Years of experience}, \; X_3 = \begin{cases} 1 & \text{Male} \\ 0 & \text{Female} \end{cases}$$
   i) Interpret each of the β's in the multiple regression model.
   ii) Fit a multiple regression model for the dataset.
   iii) Draw a residual plot to check the model adequacy.
   iv) Is there sufficient evidence to indicate that gender is an important predictor of hiring status? Use $\alpha = 0.05$.

Y	0	0	1	1	0	1	0	0	0	1	0	0
X_1	6	4	6	6	4	8	4	4	6	8	4	8
X_2	2	6	0	3	1	3	2	4	1	9	2	5
X_3	0	1	1	1	0	0	1	0	0	0	1	0