BIOS663 Homework 2
Due Wednesday, Feb 20 in class.

1. Consider a simple linear regression $Y = X\beta + \epsilon$ with an intercept and one predictor based on a sample of size 4. Or specifically,

$$\begin{pmatrix} 0.5 \\ -0.5 \\ 0.3 \\ 1.2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0.5 \\ 1 & 2 \end{pmatrix} + \epsilon. \tag{1}$$

   Calculate $(X'X)^{-1}$, $X'Y$, $\hat{\beta}$, $\hat{y}$ and $\hat{\epsilon}$ by hand.

2. Consider the model $\mathbf{y} = \beta_0 + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \beta_3\mathbf{x}_3 + \beta_4\mathbf{x}_4 + \boldsymbol{\varepsilon}$. Give the appropriate $\mathbf{C}$ and $\boldsymbol{\theta}_0$ for testing the following hypotheses.

   (a) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$

   (b) $H_0 : \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \beta_2 + 2 \\ \beta_4 \end{pmatrix}$

   (c) $H_0 : \begin{pmatrix} \beta_1 - 2\beta_2 \\ \beta_1 + 2\beta_2 \end{pmatrix} = \begin{pmatrix} 4\beta_3 \\ -6 \end{pmatrix}$

3. Consider the model $\mathbf{y}_{5\times 1} = \mathbf{X}_{5\times 3}\boldsymbol{\beta}_{3\times 1} + \boldsymbol{\varepsilon}_{5\times 1}$, where

$$\mathbf{y} = \begin{bmatrix} 5 \\ 2 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{J} & \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}, \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix},$$

   with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I})$.

   (a) Show as rigorously as possible whether $\theta_1 = \beta_2$ is estimable.

   (b) Show as rigorously as possible whether $\boldsymbol{\theta}_2 = \begin{pmatrix} \beta_0 + \beta_1 \\ \beta_0 - \beta_2 \end{pmatrix}$ is testable.

4. A group of subjects was recruited to a weight loss study in a medical center. The data consist of their weights (y = WGHT), average daily exercise times (x = TIME). One of the objectives in this study is to investigate the effect of TIME on weight loss.

   (a) A partial ANOVA table for estimating WGHT from TIME is given below. Complete the table.

```
Dependent Variable: WGHT

                              Sum of
Source            DF         Squares     Mean Square    F Value     Pr > F

Model             1       2624.670184

Error             96      1827.099916

Corrected Total 97        4451.7701
```

(b) State the model assumptions based on which the ANOVA table was computed.

(c) Is average daily exercise time a significant predictor for predicting weight loss? State your answers in terms of the model from the previous questions and the statistical test of hypothesis.

5. An investigator studied the ozone levels in the South Coast Air Basin of California for the years 1976-1991. He believes that the number of days the ozone levels exceeded 0.20 ppm (the response) depends on the seasonal meteorological index, which is the seasonal average temperature in degrees Celsius (the predictor). The data *hw2.dat*, is provided on Sakai.

   (a) Fit a regression model with the number of high ozone days as the response and the meteorological index as a covariate, and provide estimates of β_0, β_1, their standard errors, and their interpretations.

   (b) Are all of the β's estimable? Why or why not?

   (c) Report a test of the hypothesis that the number of high ozone days is associated with the meteorological index.

   (d) Using the framework of the linear model, report an $\alpha = 0.05$ test of the hypothesis that a 1 degree increase in average temperature is associated with a 12 day increase in the number of days the ozone levels exceed 0.20 ppm.

   (e) Calculate the 95% confidence interval and prediction interval for the expected number of days the ozone level exceeded 0.2 ppm when the seasonal meteorological index is 16.