Homework 3
Due Tuesday, October 1

Show all your work. Please use R for Problems 4 to 6.

1. Suppose that we are given a regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $E(\mathbf{e}) = \mathbf{0}$ and $\text{Var}(\mathbf{e}) = \sigma^2\mathbf{I}$.

   (a) Show that
   $$\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}$$

   (b) Use part (a) to show that
   $$\sum \hat{e}_i = 0$$
   (Hint: Note that $\sum \hat{e}_i = \hat{\mathbf{e}}'\mathbf{1}$, where $\mathbf{1}$ is an $n \times 1$ vector of 1's).

2. Consider a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. If we have that $E(\mathbf{e}) = \mathbf{0}$ and $\text{Var}(\mathbf{e}) = \mathbf{V}$, where $\mathbf{V}$ is an arbitrary $n \times n$ covariance matrix, find $\text{Var}(\hat{\boldsymbol{\beta}})$.

3. Suppose that we have a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$.

   (a) Show that
   $$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$
   (HINT: Subtract and add $\mathbf{X}\hat{\boldsymbol{\beta}}$ inside the $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ terms on the left hand side).

   (b) For which value of $\boldsymbol{\beta}$ is $(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$ minimized? (no need to do the differentiation; just look at the right hand side).

4. Consider the `Forbes` dataset from lecture. Use R matrix computation (similar to Example 2 of lecture) to find and verify the values from Example 3 of lecture.

5. Refer to HW 1, Problem 5. Use R (following both Example 1 and Example 2 of lecture) to compute and verify the values in Problem 5 of both HW 1 and HW 2.

6. Consider the dataset `tombstone.txt` (on the class website). The first column is the Mean SO2 Concentration (X), and the second column is the Marble Tombstone Mean Surface Recession Rate (Y). We are interested in the effect of SO2 concentration on the surface recession rate. Notice that there are no column names on the data file.

   (a) Perform the regression of Y on X, and report the estimates for β_0 and β_1, their standard errors, the value of the coefficient of determination (the R^2 and the adjusted R^2), and the residual standard error, $\hat{\sigma}$. Perform the t-test for $H_0 : \beta_0 = 0$ and $H_0 : \beta_1 = 0$ (at the level $\alpha = 0.01$).

   (b) Obtain a prediction and 90% prediction interval at $X = 100$.

   Again, follow both Example 1 and Example 2 of lecture.