STAT 3008
Homework 2
Due date: 5pm, Oct 15(Wednesday). Assignment Box at LSB 1/F.

Datasets are available at the website of the textbook: Applied Linear Regression, 3rd edition by Weisberg.

1. Show that for simple linear regression,

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

2. Given observations (x_i, y_i) for $i = 1, 2, ...n$, consider the regression model

$$y = \beta + \beta x + e\,, \qquad e \sim N(0, \sigma^2)$$

(note that they share the same β).

   i) Find the estimators for β and σ^2, i.e. $\hat{\beta}$ and $\hat{\sigma}^2$.
   ii) Find the mean and variance of $\hat{\beta}$.
   iii) Find the mean of $\hat{\sigma}^2$.
   iv) What can you say about $\hat{\beta}$ when $n \to \infty$? (Hint: What's the variance of $\hat{\beta}$ when $n \to \infty$?)
   v) Does the fitted regression line pass through $(\bar{x}, \bar{y})$? Is the sum of residuals equals zero?

3. Use the dataset **htwt.txt**, consider simple linear regression with Wt as dependent variable (response) and Ht as independent variable (predictor).

   i) Test the effect of Ht on Wt using a t-test.
   ii) Test the effect of Ht on Wt using an ANOVA table with F-test.
   iii) Find the 95% confidence interval for $E(Wt|Ht = 160)$. Draw it.
   iv) Find the 90% simultaneous confidence band for $E(Wt|Ht)$. Draw it.
   v) Find the 99% prediction interval for a new observations at $Ht = 160$.
   vi) Draw a residual plot. Is the regression a good fit?

4. Let

$$X = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}. \quad E(Y) = \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix}. \quad Var(Y) = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(i) Find $X(X'X)^{-1}X'$, $(X'X)^{-1}X'Y$, $X(X'X)^{-1}X'Y$ and $\bar{Y}$, in terms of Y_1, Y_2, Y_3.

(ii) Base on (i), is "Taking sample mean" the same as "Regression using constant term only"?

(iii) Let $H = X(X'X)^{-1}X'$. Express $Y'(1-H)Y$ in terms of Y_1, Y_2, Y_3. Then find $E(Y'(1-H)Y)$.

(iv) Using the fact that $E(X'AX) = E[tr(X'AX)] = tr[AE(XX')]$, find $E(Y'(I-H)Y)$.

(v) Repeat (iii) and (iv) using $Var(Y) = I_3$.

5. 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.

(i) Using matrix methods find a)$(X'X)^{-1}$, b) $\hat{\beta}$, c)$\hat{\mathbf{e}}$, d)$H = X(X'X)^{-1}X'$, e) SYY, f) SSReg, g)RSS, h)$\hat{\sigma}^2$,i)$\hat{Y}$, j) R^2.

(ii) Find the variance-covariance matrix of $\hat{\beta}$ in terms of σ. Note that σ is an unknown parameter. How do you estimate this matrix from the data?

(iii) What is the variance-covariance matrix of β (not $\hat{\beta}$)?

(iv) Find the mean of $\hat{\sigma}^2$ in terms of σ.

(v) Estimate $Var(\hat{Y})$. Are Y_is independent? Are $\hat{Y}_i$s independent?

(vi) Write an ANOVA table to test the effect of the regression.

(vii) At the point $(X1, X2) = (12, -2)$, estimate $E(Y|X1, X2)$ and

a) 90% confidence interval for the fitted value of $(X1, X2)$.

b) 95% prediction interval for a new observation of $(X1, X2)$.

Will you trust the estimates? Why?