Stat 252.01 Winter 2005
Assignment #9 Solutions

(11.56) For the quadratic model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ given in the problem, we know that the least squares estimators $\hat{\beta}_i$ can be found by solving the matrix equation

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

where

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \\ 1 & x_6 & x_6^2 \\ 1 & x_7 & x_7^2 \end{bmatrix} \qquad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

From the data given in the problem, we have

$$\mathbf{Y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}.$$

Hence,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \end{bmatrix} \cdot \begin{bmatrix} 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 7 & 0 & 28 \\ 0 & 28 & 0 \\ 28 & 0 & 196 \end{bmatrix}$$

so that

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/3 & 0 & -1/21 \\ 0 & 1/28 & 0 \\ -1/21 & 0 & 1/84 \end{bmatrix}.$$

Furthermore,

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -4 \\ 8 \end{bmatrix}.$$

Finally, we can conclude that

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1/3 & 0 & -1/21 \\ 0 & 1/28 & 0 \\ -1/21 & 0 & 1/84 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -4 \\ 8 \end{bmatrix} = \begin{bmatrix} -15/21 \\ -1/7 \\ 1/7 \end{bmatrix}.$$

Hence, the quadratic regression fitted to the data is given by

$$\hat{y} = (-15/21) - (1/7)\,x + (1/7)\,x^2.$$

2. (b) Referring to Problem #5 on Midterm #2, we see that

$$\mathrm{SSE}(\beta_0,\beta_1) = \mathbb{E}[(Y-\hat{Y})^2] = (\mu_y - \beta_0 - \beta_1\mu_x)^2 + \sigma_y^2 + \beta_1^2\sigma_x^2 - 2\beta_1\sigma_{xy},$$

from which we compute

$$\frac{\partial}{\partial\beta_0}\mathrm{SSE}(\beta_0,\beta_1) = -2(\mu_y-\beta_0-\beta_1\mu_x) \text{ and } \frac{\partial}{\partial\beta_1}\mathrm{SSE}(\beta_0,\beta_1) = -2\mu_x(\mu_y-\beta_0-\beta_1\mu_x)+2\beta_1\sigma_x^2-2\sigma_{xy}.$$

Furthermore,

$$\frac{\partial^2}{\partial\beta_0^2}\mathrm{SSE}(\beta_0,\beta_1) = 2, \quad \frac{\partial^2}{\partial\beta_1^2}\mathrm{SSE}(\beta_0,\beta_1) = 2\mu_x^2 + 2\sigma_x^2, \quad \text{and}$$

$$\frac{\partial^2}{\partial\beta_0\beta_1}\mathrm{SSE}(\beta_0,\beta_1) = \frac{\partial^2}{\partial\beta_1\beta_0}\mathrm{SSE}(\beta_0,\beta_1) = 2\mu_x.$$

Since we have shown already that the partial derivatives equal zero when

$$\beta_0 = \mu_y - \beta_1\mu_x$$

and

$$\beta_1 = \frac{\sigma_{xy}}{\sigma_x^2},$$

in order for these to be the minimizing values of $\mathrm{SSE}(\beta_0,\beta_1)$ we must check the second derivative test. (A statement of this result can be found on the handout dated March 10.) Since,

$$\frac{\partial^2}{\partial\beta_0^2}\mathrm{SSE}(\beta_0,\beta_1) = 2 > 0$$

and since

$$\frac{\partial^2}{\partial\beta_0^2}\mathrm{SSE}(\beta_0,\beta_1)\cdot\frac{\partial^2}{\partial\beta_1^2}\mathrm{SSE}(\beta_0,\beta_1) - \left(\frac{\partial^2}{\partial\beta_0\beta_1}\mathrm{SSE}(\beta_0,\beta_1)\right)^2 = 2\cdot(2\mu_x^2+2\sigma_x^2) - (2\mu_x)^2 = 4\sigma_x^2 > 0,$$

we conclude by the second derivative test that β_0 and β_1 as given do minimize SSE.

2. (c) Consider the modified linear regression model described by Problem #5 on Midterm #2 which outlines a method of predicting the random variable Y based on information contained in another random variable X. In particular, the predictor of Y, called $\hat{Y}$, is assumed to be a linear function of X so that $\hat{Y} = \beta_0 + \beta_1 X$. The problem discusses one way of selecting the coefficients β_i, namely by choosing them to minimize the mean square error $\mathbb{E}[(Y-\hat{Y})^2]$. As shown on the Midterm, and in the exercise above, the minimizing values of β_i are given by

$$\beta_0 = \mu_y - \beta_1\mu_x \quad \text{and} \quad \beta_1 = \frac{\sigma_{xy}}{\sigma_x^2}.$$

This implies that

$$\hat{Y} = \mu_y - \frac{\sigma_{xy}\mu_x}{\sigma_x^2} + \frac{\sigma_{xy}}{\sigma_x^2}X.$$

One thing to immediately notice is that the best linear predictor of Y (in the minimum mean square error sense) depends not only on the random variable X, but on some ancillary information about Y, namely its mean, μ_y; its variance, σ_y^2; and its covariance with X, σ_{xy}. Suppose, however, that $\mathrm{Cov}(X, Y) = 0$. Then, in the notation of the problem at hand, $\sigma_{xy} = 0$ so that

$$\beta_1 = \frac{\sigma_{xy}}{\sigma_x^2} = 0 \quad \text{and} \quad \beta_0 = \mu_y - \beta_1 \mu_x = \mu_y,$$

and

$$\hat{Y} = \mu_y.$$

This says that the best linear predictor (in the mean square sense) of Y is now simply μ_y, the mean of Y. Notice, in particular, that knowledge of X has no bearing on prediction of Y. It is in this sense that uncorrelatedness is *similar* to independence. (Although, of course, dependent random variables may be uncorrelated.)

(11.81) Suppose that

$$\mathbf{Y} = [Y_1 \, Y_2 \, \cdots \, Y_n] \quad \text{and} \quad \mathbb{1} = [1 \, 1 \, \cdots \, 1]$$

are both $1 \times n$ vectors. Then, we can write

$$\overline{Y} = \frac{1}{n} \mathbb{1}' \mathbf{Y} = [1/n \, 1/n \, \cdots \, 1/n] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}.$$

In matrix form, we are interested in the equation

$$Y = \boldsymbol{x}' \hat{\boldsymbol{\beta}}$$

where

$$\boldsymbol{x}' = [x_1 \, x_2 \, \cdots \, x_k] \quad \text{and} \quad \hat{\boldsymbol{\beta}}' = [\hat{\beta}_1 \, \hat{\beta}_1 \, \cdots \, \hat{\beta}_k].$$

Suppose that $Y = \overline{Y}$ so that

$$\overline{Y} = \boldsymbol{x}' \hat{\boldsymbol{\beta}} = \boldsymbol{x}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}.$$

This implies

$$\frac{1}{n} \mathbb{1}' \mathbf{Y} = \boldsymbol{x}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$$

so that

$$\frac{1}{n} \mathbb{1}' = \boldsymbol{x}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}',$$

or

$$\frac{1}{n} \mathbb{1}' \mathbf{X} = \boldsymbol{x}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{X} = \boldsymbol{x}'.$$

Hence, we conclude that

$$\boldsymbol{x}' = \frac{1}{n} \mathbb{1}' \mathbf{X} = [1 \, \overline{x}_1 \, \overline{x}_2 \, \cdots \, \overline{x}_k].$$

In other words, the point $(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k, \overline{Y})$ satisfies the equation $Y = \boldsymbol{x}' \hat{\boldsymbol{\beta}}$ so that the least squares prediction line must pass through this point.