Applied Math 166 Problem Set 9 4/25/03

11.66 From Exercise 11.27, the complete model is $Y = \beta_0 + \beta_1 x + \epsilon$ with $\text{SSE}_c = 6.452$ (8 d.f.) The reduced model is $Y = \beta_0 + \epsilon$ with $\widehat{\beta}_0 = \overline{Y}$ and $\text{SSE}_R = \sum(Y_i - \widehat{Y})^2 = \sum(Y_i - \overline{Y})^2 = 99.149$ (9 d.f.). The test statistic for testing H_0: $\beta_1 = 0$ is

$$F = \frac{S_3^2}{S_2^2} = \frac{\frac{\text{SSE}_R - \text{SSE}_c}{9-8}}{\frac{\text{SSE}_c}{8}} = \frac{92.697}{\frac{6.452}{8}} = 114.94$$

The rejection region with $\alpha = .05$ is $F_{1,8} = 5.32$, and H_0 is rejected. There is evidence to indicate that $\beta_1 \neq 0$.

11.67 First, note that with $k = 1$, $\text{SSE}_R = S_{yy}$. Then, $g = 0$, and $\text{SSE}_c = S_{yy} - \widehat{\beta}_1 S_{xy}$.

$$F = \frac{(\text{SSE}_R - \text{SSE}_c)(k-g)}{\frac{\text{SSE}_c}{n-(k+1)}} = \frac{S_{yy} - S_{yy} + \widehat{\beta}_1 S_{xy}}{\frac{\text{SSE}_c}{n-2}} = \frac{\widehat{\beta}_1 S_{xy}}{\frac{\text{SSE}_c}{n-2}} = \frac{\frac{\widehat{\beta}_1 S_{xy}}{S_{xy}}}{\frac{s^2}{S_{xx}}} = \frac{\widehat{\beta}_1^2}{\frac{s^2}{S_{xx}}}$$
$$= t^2$$

In Exercise 11.66, $F = 114.94$, and in Exercise 11.27, $t = -10.72$. Thus, $t^2 = 1$ which equals F (except for roundoff error).

11.68 **a.** Refer to Exercise 11.57. For the complete model, $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$, $\text{SSE}_c = 245.69$ (see Exercise 11.60) with 7 degrees of freedom. For the reduced model, $Y = \beta_0 + \beta_1 x + \epsilon$,

$$\text{SSE}_R = \boldsymbol{Y'Y} - \beta_1 \boldsymbol{X'Y} = 10.208.67 - [29.99 \quad 1.3879] \begin{bmatrix} 299.9 \\ 458.3 \end{bmatrix} = 578.6$$

with 8 d.f. The test statistic for testing H_0: $\beta_2 = 0$ is

$$F = \frac{\frac{\text{SSE}_R - \text{SSE}_c}{8-7}}{\frac{\text{SSE}_c}{7}} = \frac{332.91}{\frac{245.69}{7}} = 9.49$$

The rejection region with $\alpha = .05$ is $F > F_{1,7} = 5.59$, and H_0 is rejected. There is evidence that $\beta_2 \neq 0$. These results do agree with the results of problem 11.60.

b. For the reduced model, $Y = \beta_0 + \epsilon$, $\text{SSE}_R = \sum(y_i - \overline{y})^2 = 1214.669$ with 9 d.f. Then

$$F = \frac{\frac{\text{SSE}_R - \text{SSE}_c}{9-7}}{\frac{\text{SSE}_c}{7}} = \frac{\frac{968.98}{2}}{\frac{245.69}{7}} = 13.80$$

The rejection region with $\alpha = .05$ is $F > F_{2,7} = 4.74$, and the null hypothesis, H_0: $\beta_1 = \beta_2 = 0$, is rejected.

11.69 The hypothesis of interest is H_0: $\beta_1 - \beta_4 = 0$. For the complete model,

$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$, $\quad \text{SSE}_c = 98.8125$ with 11 d.f.

For the reduced model, the 2nd and 5th columns of the $\boldsymbol{X}$ matrix are deleted, so that

$$(\boldsymbol{X'X})^{-1} = \begin{bmatrix} \frac{1}{16} & 0 & 0 \\ 0 & \frac{1}{16} & 0 \\ 0 & 0 & \frac{1}{16} \end{bmatrix} \qquad \boldsymbol{X'Y} = \begin{bmatrix} 338 \\ -19.4 \\ -2.6 \end{bmatrix}$$

Then

$$\text{SSE} = \boldsymbol{Y'Y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X'Y} = 7446.52 - 7164.195 = 282.325$$

with 13 d.f. The test statistic is

$$F = \frac{\frac{\text{SSE}_R - \text{SSE}_c}{2}}{\frac{\text{SSE}_c}{11}} = 10.21$$

The rejection region with $\alpha = .05$ is $F > F_{2,11} = 3.98$, and H_0 is rejected. There is reason to believe that either T_1 or T_2 or both affect the yield.

11.70 We need to carry out an F test with test statistic

$$F = \frac{\frac{\text{SSE}_R - \text{SSE}_c}{3}}{\frac{\text{SSE}_c}{24-6}} = \frac{\frac{465.134 - 152.177}{3}}{\frac{152.177}{18}} = 12.34$$

Refer to Table 7. We have $12.34 > 5.92 = F_{.005}$, thus the p-value $< .005$.

11.80 The F statistic is

$$F = \frac{\frac{\text{SSE}_1 \quad \text{SSE}_2}{2}}{\frac{\text{SSE}_2}{200-5}} = \frac{\frac{795.23 \quad 783.9}{2}}{\frac{783.9}{195}} = 1.41$$

The critical value is $F_{.05} = 3.00 > 1.41$, so we do not reject H_0: salary is not dependent on sex.

11.81 Let

$$\boldsymbol{Y}' = [\,Y_1 \quad Y_2 \quad Y_3 \quad \cdots \quad Y_n\,] \qquad \text{and} \qquad \mathbf{1}' = [\,1 \quad 1 \quad 1 \quad \cdots \quad 1\,]$$

be two $1 \times n$ vectors. Then we can write

$$\overline{Y} = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \frac{1}{n}\,\mathbf{1}'\boldsymbol{Y}$$

In matrix form, the equation of interest is $Y = \boldsymbol{x}'\widehat{\boldsymbol{\beta}}$, where

$$\boldsymbol{x}' = [\,1 \quad x_1 \quad x_2 \quad \cdots \quad x_k\,]$$

and

$$\widehat{\boldsymbol{\beta}}' = [\,\widehat{\beta}_0 \quad \widehat{\beta}_1 \quad \cdots \quad \widehat{\beta}_k\,].$$

Suppose $Y = \overline{Y}$. Then

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

implying

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

Which must then imply

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

or

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

or

$$\frac{1}{n}\,\mathbf{1}'\boldsymbol{X} = \boldsymbol{x}'.$$

That is

$$\boldsymbol{x}' = [\,1 \quad \overline{x}_1 \quad \overline{x}_2 \quad \cdots \quad \overline{x}_k\,].$$

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

3

11.82a. Use the coding

$$X_1^* = \frac{X_1 - 65}{15}$$

and

$$X_2^* = \frac{X_2 - 200}{100}.$$

Then $X_1 = 50, 80$ correspond to $X_1^* = -1, 1$, while $X_2 = 100, 200, 300$ correspond to $X_2^* = -1, 0, 1$. Now

$$\boldsymbol{Y} = \begin{bmatrix} 21 \\ 23 \\ 26 \\ 22 \\ 23 \\ 28 \end{bmatrix} \quad \boldsymbol{X}^* = \begin{array}{cccc} x_0 & x_1 & x_2 & x_2^2 \end{array} \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \boldsymbol{X}^{*\prime}\boldsymbol{X}^* = \begin{bmatrix} 6 & 0 & 0 & 4 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 4 & 0 & 0 & 4 \end{bmatrix}$$

$$(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{3}{4} \end{bmatrix} \quad \boldsymbol{X}^{*\prime}\boldsymbol{Y} = \begin{bmatrix} 143 \\ 3 \\ 11 \\ 97 \end{bmatrix} \quad \widehat{\boldsymbol{\beta}}^* = \begin{bmatrix} 23 \\ .5 \\ 2.75 \\ 1.25 \end{bmatrix}$$

Finally, the least squares equation is

$$\begin{aligned} \widehat{y} &= 23 + .5\left(\tfrac{x_1-65}{15}\right) + 2.75\left(\tfrac{x_2-200}{100}\right) + 1.25\left(\tfrac{x_2-200}{100}\right)^2 \\ &= 20.33 + .0333x_1 - .0225x_2 + .000125x_2^2 \end{aligned}$$

b. The hypothesis of interest, H_0: $\beta_3 = 0$, is equivalent to a test of H_0: $\beta_3^* = 0$ since $\beta_3 = \left(\frac{1}{100}\right)^2\beta_3^*$. Using the coded calculations, then, we have

$$\text{SSE} = \boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}'\boldsymbol{Y} = 3443 - 3442 = 1$$

and

$$s^2 = \frac{\text{SSE}}{n-4} = .5.$$

The test statistic is

$$t = \frac{\widehat{\beta}_3^*}{\sqrt{c_{33}s^2}} = \frac{1.25}{\sqrt{\left(\frac{1}{2}\right)\left(\frac{3}{4}\right)}} = 2.042$$

The rejection region with $\alpha = .05$ is $|t| > 4.303$, and H_0 is not rejected. The quadratic temperature effect is not significant.

c. In order to test H_0: $\beta_2 = \beta_3 = 0$, or, equivalently, H_0: $\beta_2^* = \beta_3^* = 0$, complete and reduced models are fitted. For the complete model, $\text{SSE}_2 = 1$ with 2 d.f. [from part **b**]. For the reduced model, columns 3 and 4 of the $\boldsymbol{X}^*$ matrix are omitted and

$$(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1} = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{6} \end{bmatrix}$$

$$\boldsymbol{X}^*\boldsymbol{Y} = \begin{bmatrix} 143 \\ 3 \end{bmatrix}$$

so that

$$\text{SSE}_1 = 3443 - [23.8333 \quad .5]\begin{bmatrix} 143 \\ 3 \end{bmatrix} = 33.33.$$

with 4 d.f. The test statistic is then

$$F = \frac{\frac{\text{SSE}_1 - \text{SSE}_2}{2}}{\frac{\text{SSE}_2}{2}} = \frac{\frac{33.33-1}{2}}{\frac{1}{2}} = 32.33$$

The rejection region with $\alpha = .05$ is $F > F_{2,2} = 19.00$, and the null hypothesis is rejected. Temperature does affect yield.