# Math 644, Fall 2012

## Solution for Homework 4

**Problem 1.**

### 6.1

a. $$X = \begin{bmatrix} 1 & X_{11} & X_{11}X_{12} \\ 1 & X_{21} & X_{21}X_{22} \\ 1 & X_{31} & X_{31}X_{32} \\ 1 & X_{41} & X_{41}X_{42} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

b. $$X = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ 1 & X_{31} & X_{32} \\ 1 & X_{41} & X_{42} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

### 6.2

a. $$X = \begin{bmatrix} X_{11} & X_{12} & X_{11}^2 \\ X_{21} & X_{22} & X_{21}^2 \\ X_{31} & X_{32} & X_{31}^2 \\ X_{41} & X_{42} & X_{41}^2 \\ X_{51} & X_{52} & X_{51}^2 \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

b. $$X = \begin{bmatrix} 1 & X_{11} & \log_{10} X_{12} \\ 1 & X_{21} & \log_{10} X_{22} \\ 1 & X_{31} & \log_{10} X_{32} \\ 1 & X_{41} & \log_{10} X_{42} \\ 1 & X_{51} & \log_{10} X_{52} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

**Problem 2. See the following R code.**

```
con = c(1, 1, 1, 1, 1, 1)
X1 = c(7, 4, 16, 3, 21, 8)
X2 = c(33, 41, 7, 49, 5, 31)
Y = c(42, 33, 75, 28, 91, 55)

X = matrix(c(con, X1, X2), 6, 3);

XX = t(X) %*% X
XY = t(X) %*% Y

b = solve(XX) %*% XY
b

Yhat = X %*% b;
e = Y - Yhat
e

SSR = sum( (Yhat - mean(Y))^2 )
SSR

SSE = sum( e^2  )
MSE = SSE/(6-3)

s2b = diag(solve(XX))*MSE
s2b

Xnew = matrix(c(1, 10, 30), 3, 1);
Ynewhat = t(Xnew) %*% b
Ynewhat

s2Ynewhat = MSE*(1 + t(Xnew) %*%  solve(XX) %*% Xnew)
s2Ynewhat
```

**Problem 3.**

A student stated: "Adding predictor variables to a regression model can never reduce $R^2$, so we should include all available predictor variables in the model." Comment.

Bigger $R^2$, means the fitting is better. Better fitting does not imply better model.

**Problem 4. (a)**

$$F = (364.2/1)/(4248.8/42) = 3.600169$$

Since $3.600169 < F(0.975, 1, 42) = 5.4239$, $X_3$ can be dropped. The p-value is 0.06468

**(b)**

Full model

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Reduced model

$$y = \beta_0 - X_1 + \beta_3 X_3 + \varepsilon$$

SSE(F) = 4248.8, SSE(R) = 4427.7

$$F^* = \frac{(SSE(R) - SSE(F))/2}{SSE(F)/42} = \frac{(4427.7 - 4248.8)/2}{4248.8/42} = 0.8842 < F(0.975, 2, 42) = 4.0510$$

Thus, $\beta_1 = -1.0$ and $\beta_2 = 0$ can be accepted.

**Problem 5.**

The following regression model is being considered in a water resources study: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \beta_4 \sqrt{X_{i3}} + \varepsilon_i$. State the reduced model for testing whether or not: (1) $\beta_3 = \beta_4 = 0$, (2) $\beta_3 = 0$, (3) $\beta_1 = \beta_2 = 5$, (4) $\beta_4 = 7$.

(1)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

(2)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_4 \sqrt{X_{i3}} + \varepsilon_i$$

(3)

$$Y_i = \beta_0 + 5X_{i1} + 5X_{i2} + \beta_3 X_{i1} X_{i2} + \beta_4 \sqrt{X_{i3}} + \varepsilon_i$$

(4)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + 7\sqrt{X_{i3}} + \varepsilon_i$$