Homework 4 Solutions:
Chapter 3: #15
Chapter 5: #, 5, 8ab, 17a; Chapter 6: #2, 25; Chapter 8: 21a; Show that H is idempotent

Assigned Mon, Oct 21:

3.15 **a.** Fit a linear regression function: $\hat{Y} = 2.5753 - 0.3240$ X, found using these R commands:

```
solcon <- read.table('CH03PR15.txt',col.names=c('Y','X'))
LR = lm(Y~X,data=solcon)
summary(LR)
```

b. Perform the lack of fit test using $\alpha = .025$. Results show that there is lack of fit. The hypotheses are given below in words and in symbols; either is okay:
H_0: Linear regression model is as good as specifying separate means μ_j for the 5 time periods.
H_a: H_0 is not true; the regression model does not fit as well as allowing separate means
In symbols:
H_0: $Y_{ij} = \beta_0 + \beta_1 X_j + \varepsilon_{ij}$
H_a: $Y_{ij} = \mu_j + \varepsilon_{ij}$

<u>R commands</u>:

```
Reduced <- lm(Y~X,data=solcon)
Full <- lm(Y~0+as.factor(X),data=solcon)
anova(Reduced,Full)
```

<u>R Output</u>:

```
Analysis of Variance Table

Model 1: Y ~ X
Model 2: Y ~ 0 + as.factor(X)
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     13 2.9247
2     10 0.1574  3    2.7673 58.603 1.194e-06 ***
```

The p-value is shown in bold. So the ANOVA shows we should reject the null hypothesis that the reduced model fits as well as the full model at the level 0.025. There is lack of fit in the fitted linear regression. Allowing separate means (treating X as a categorical variable) models the relationship between Y and X better than does linear regression.
c. No, the test does not give any information about what regression function is appropriate.

Assigned Wed, October 23

5.5 You can do this using R (see handout on website “Matrices in R”), or do them by hand. The data set is small so it’s not that difficult to do them by hand. Here are the formulas and results.

(1) $\mathbf{Y'Y} = \sum_{i=1}^{n} Y_i^2 = 1259$

(2) $\mathbf{X'X} = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix} = \begin{bmatrix} 6 & 17 \\ 17 & 55 \end{bmatrix}$

(3) $\mathbf{X'Y} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix} = \begin{bmatrix} 81 \\ 261 \end{bmatrix}$

5.8 **a.** The columns are linearly dependent (not linearly independent). Note that $\mathbf{C}_1 = \mathbf{C}_2 + \mathbf{C}_3$
b. The rank is 2, the number of columns that are linearly independent.

5.17 **a.** $\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$

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

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

6.25 Use $Y_i' = Y_i - 4 X_{i2}$ and then fit the model response variable Y_i' and explanatory variables X_1 and X_3.

8.21 **a.** We will write each response function in the form of intercept + slope×X.
Hard hat: $E\{Y\} = (\beta_0 + \beta_2) + \beta_1 X_1$
Bump cap: $E\{Y\} = (\beta_0 + \beta_3) + \beta_1 X_1$
None: $E\{Y\} = \beta_0 + \beta_1 X_1$

Use matrix algebra to show that the hat matrix H is idempotent

The hat matrix is $\mathbf{H} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$, so $\mathbf{HH} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1}\mathbf{X'} = \mathbf{XI}(\mathbf{X'X})^{-1}\mathbf{X'} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'} = \mathbf{H}$.