STAT 8004 – Statistical Methods II
Spring 2015
Homework Assignment 2 (Due on 2/12/2015 before the end of the day)

- Reading assignment
  - R&S Chapter 7.
  - Re-parametrization part is in R&S Chapter 12.
  - There are various resources with fitting linear models with R, e.g. the suggested text book of the course and webs such as `http://data.princeton.edu/R/linearModels.html`.
- The following exercises are to be collected. Please upload your homework to the Blackboard. Following the requirement of STAT 8003, please typeset your homework with Latex and upload both the pdf and latex files.

1. Suppose that we are working under Gauss-Markov model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$. Let $\hat{\mathbf{Y}}$ be the ordinary least square estimator of $\mathbf{Y}$.

   (a) Show that $\hat{\mathbf{Y}}$ and $\mathbf{Y} - \hat{\mathbf{Y}}$ are uncorrelated.

   (b) Show that

$$E\{(\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}})\} = \sigma^2\{n - \mathrm{rank}(\mathbf{X})\}.$$

   You may use Theorem 5.2a of R&S.

2. Consider the one-way ANOVA model $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ for the jth individual of the ith group. Suppose there are 4 treatments (groups) and the sample sizes are respectively 2,1,1,2 for treatments. Now suppose that $\mathbf{Y} = (y_{11}, y_{12}, y_{21}, y_{31}, y_{41}, y_{42})^T = (2, 1, 4, 6, 3, 5)^T$ contains the observations.

   Use R and weighted generalized least squares to find appropriate estimate for

$$E(\mathbf{Y}) \text{ and } \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{\beta}$$

   in the Aiken model with $\mathrm{var}(\boldsymbol{\varepsilon}) = \mathbf{V}$ for two cases where

   (a) $\mathbf{V} = \mathbf{V}_1 = \mathrm{diag}(1, 9, 9, 1, 1, 9)$ and

   (b)

$$\mathbf{V} = \mathbf{V}_2 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 9 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 9 \end{pmatrix}$$

.

3. The `lm` function in R allows one to do weighted least squares with the form $\sum w_i(y_i - \hat{y}_i)^2$ for positive weights w_i. For $\mathbf{V}_1$ in the last question, find the BLUEs of the 4 cell means using `lm` and an appropriate vector of weights.

4. By running

```
library(MASS)
data(Boston)
```

will load the Boston housing data into R. Use `?Boston` to see the information on the variables. Now create two matrices $\mathbf{Y}$ and $\mathbf{X}$ that will be used to fit a regression model to some of these data.

```
Y=as.matrix(Boston$medv)
X=as.matrix(Boston[,c('crim','nox','rm','age','dis')])
X=cbind(rep(1,dim(Boston)[1]),X)
```

(a) Make a scatterplot matrix for $y, x_1, \ldots, x_5$. If you had to guess based on this plot, which single predictor do you think is probably the best predictor of Price? Do you see any evidence of multicollinearity (correlation among the predictors) in this graphic?

(b) Use `qr()` function to find the rank of $\mathbf{X}$.

(c) Use R matrix operations on the $\mathbf{X}$ matrix and $\mathbf{Y}$ vector to find the estimated regression coefficient vector $\hat{\boldsymbol{\beta}}$, the estimated mean vector $\hat{\mathbf{Y}}$, and the vector of residuals $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$.

(d) Plot the residuals against the fitted means.

(e) Create a normal plot from the values in the residual vector.

(f) Compute the sum of squared residuals and the corresponding estimate of σ^2

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}})}{n - \text{rank}(\mathbf{X})}.$$

(g) Call the `lm` function in R and confirm your answers, and note that `?lm` gives you various information such as the outputs of the function.

```
m1=lm(medv~crim+nox+rm+age+dis,data=Boston)
```