Stat 511 HW#2 Spring 2008

1. In class Vardeman argued that hypotheses of the form $\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ can be written as $\mathrm{H}_0 : \mathrm{E}\,\mathbf{Y} \in C(\mathbf{X}_0)$ for $\mathbf{X}_0$ a suitable matrix (and $C(\mathbf{X}_0) \subset C(\mathbf{X})$). Let's investigate this notion in the context of Problem 10 of Homework 1. Consider

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 1 & -1 & 0 & 0 & .5 & .5 & -.5 & -.5 \end{pmatrix}$$

and the hypothesis $\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.

a) Find a matrix $\mathbf{A}$ such that $\mathbf{C} = \mathbf{AX}$.

b) Let $\mathbf{X}_0$ be the matrix consisting of the 1st, 4th and 5th columns of $\mathbf{X}$. Argue that the hypothesis under consideration is equivalent to the hypothesis $\mathrm{H}_0 : \mathrm{E}\,\mathbf{Y} \in C(\mathbf{X}_0)$. (Note: One clearly has $C(\mathbf{X}_0) \subset C(\mathbf{X})$. To show that $C(\mathbf{X}_0) \subset C(\mathbf{A}')^{\perp}$ it suffices to show that $\mathbf{P}_{\mathbf{A}'}\mathbf{X}_0 = \mathbf{0}$ and you can use R to do this. Then the dimension of $C(\mathbf{X}_0)$ is clearly 2, i.e. $\mathrm{rank}(\mathbf{X}_0) = 2$. So $C(\mathbf{X}_0)$ is a subspace of $C(\mathbf{X}) \cap C(\mathbf{A}')^{\perp}$ of dimension 2. But the dimension of $C(\mathbf{X}) \cap C(\mathbf{A}')^{\perp}$ is itself $\mathrm{rank}(\mathbf{X}) - \mathrm{rank}(\mathbf{C}) = 4 - 2 = 2$.)

c) Interpret the null hypothesis under discussion here in Stat 500 language.

2. Suppose we are operating under the (common Gauss-Markov) assumptions that $\mathrm{E}\,\boldsymbol{\varepsilon} = \mathbf{0}$ and $\mathrm{Var}\,\boldsymbol{\varepsilon} = \sigma^2 \mathbf{I}$.

a) Use fact 1. of Appendix 7.1 of the 2004 class outline to find $\mathrm{E}(\mathbf{Y} - \hat{\mathbf{Y}})$ and $\mathrm{Var}(\mathbf{Y} - \hat{\mathbf{Y}})$. (Use the fact that $\mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}$.) Then write

$$\begin{pmatrix} \hat{\mathbf{Y}} \\ \mathbf{Y} - \hat{\mathbf{Y}} \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{\mathbf{X}} \\ \mathbf{I} - \mathbf{P}_{\mathbf{X}} \end{pmatrix} \mathbf{Y}$$

and use fact 1 of Appendix 7.1 to argue that every entry of $\mathbf{Y} - \hat{\mathbf{Y}}$ is uncorrelated with every entry of $\hat{\mathbf{Y}}$.

b) Theorem 5.2.A of Rencher or Theorem 1.3.2 of Christensen say that if $\mathrm{E}\mathbf{Y} = \boldsymbol{\mu}$ and $\mathrm{Var}\,\mathbf{Y} = \boldsymbol{\Sigma}$ and $\mathbf{A}$ is a symmetric matrix of constants, then

$$\mathrm{E}\,\mathbf{Y}'\mathbf{A}\mathbf{Y} = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$$

Use this fact and argue carefully that

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

3. a) In the context of Problem 3 of HW 1 and the fake data vector used in Problem 4 of HW 1, use `R` and generalized least squares to find appropriate estimates for

$$\mathrm{E}\,\mathbf{Y} \text{ and } \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{\beta}$$

in the Aitken models with

$$\text{first } \mathbf{V}_1 = \text{diag}(1,4,4,1,1,4,4) \text{ and then } \mathbf{V}_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

(Do the necessary matrix calculations in `R`.)

b) For both of the above covariance structures, compare the (Aitken model) covariance matrices for generalized least squares estimators to (Aitken model) covariance matrices for the OLS estimators of $\mathrm{E}\,\mathbf{Y}$ and the $\mathbf{C}\boldsymbol{\beta}$ above.

4. a) The basic `lm` function in `R` allows one to automatically do weighted least squares, i.e. minimize $\sum w_i \left(y_i - \hat{y}_i \right)^2$ for positive weights w_i. For the $\mathbf{V}_1$ case of the Aitken model of Problem 3, find the BLUEs of the 4 cell means using `lm` and an appropriate vector of weights. (Type `> help(lm)` in `R` in order to get help with the syntax.)

b) The `lm.gls()` function in the `R` contributed package `MASS` allows one to do generalized least squares as described in class. For the $\mathbf{V}_2$ case of the Aitken model of Problem 3, find the BLUEs of the 4 cell means using `lm.gls`. (After loading the `MASS` package, Type `> help(lm.gls)` in order to get help with the syntax.)