Stat 511 HW#2 Spring 2008

1. In class Vardeman argued that hypotheses of the form $H_0 : C\beta = 0$ can be written as $H_0 : EY \in C(X_0)$ for $X_0$ a suitable matrix (and $C(X_0) \subseteq C(X)$). Let’s investigate this notion in the context of Problem 10 of Homework 1. Consider

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 1 & -1 & 0 & 0.5 & 0.5 & -0.5 & -0.5 \end{pmatrix}$$

and the hypothesis $H_0 : C\beta = 0$.

a) Find a matrix $A$ such that $C = AX$.

b) Let $X_0$ be the matrix consisting of the 1st, 4th and 5th columns of $X$. Argue that the hypothesis under consideration is equivalent to the hypothesis $H_0 : E\gamma \in C(X_0)$. (Note: One clearly has $C(X_0) \subseteq C(X)$. To show that $C(X_0) \subseteq C(A')^\perp$ it suffices to show that $P_A X_0 = 0$ and you can use $\mathbb{R}$ to do this. Then the dimension of $C(X_0)$ is clearly 2, i.e. $\text{rank}(X_0) = 2$. So $C(X_0)$ is a subspace of $C(X) \cap C(A')^\perp$ of dimension 2. But the dimension of $C(X) \cap C(A')^\perp$ is itself $\text{rank}(X) - \text{rank}(C) = 4 - 2 = 2$.)

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

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

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

$$\begin{pmatrix} \hat{Y} \\ Y - \hat{Y} \end{pmatrix} = \begin{pmatrix} P_X \\ I - P_X \end{pmatrix} Y$$

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

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

$$EY'AY = \text{tr}(A\Sigma) + \mu'A\mu$$

Use this fact and argue carefully that

$$E\left(Y - \hat{Y}\right)'\left(Y - \hat{Y}\right) = \sigma^2 \left(n - \text{rank}(X)\right)$$
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

\[ \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1
\end{bmatrix} \beta \]

in the Aitken models with

\[
V_1 = \text{diag}(1, 4, 4, 1, 4, 4) \quad \text{and then} \quad V_2 = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & -1 & 0 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 4 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 4
\end{bmatrix}
\]

(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 EY and the C\beta above.

4. a) The basic \texttt{lm} function in R allows one to automatically do weighted least squares, i.e. minimize \( \sum w_i (y_i - \hat{y}_i)^2 \) for positive weights \( w_i \). For the \( V_1 \) case of the Aitken model of Problem 3, find the BLUEs of the 4 cell means using \texttt{lm} and an appropriate vector of weights. (Type \texttt{> help(lm)} in R in order to get help with the syntax.)

b) The \texttt{lm.gls()} function in the R contributed package \texttt{MASS} allows one to do generalized least squares as described in class. For the \( V_2 \) case of the Aitken model of Problem 3, find the BLUEs of the 4 cell means using \texttt{lm.gls}. (After loading the \texttt{MASS} package, Type \texttt{> help(lm.gls)} in order to get help with the syntax.)