Stat 511 Exam I: Solutions and Comments Spring 2001

There may be more than one way to correctly answer a question, or several ways to describe the same answer. Not all of the possible correct, reasonable or partially correct solutions may be listed here.

1. (A) (5 points) A linear model with $E(\mathbf{Y}) = X\boldsymbol{\beta}$ and $Var(\mathbf{Y}) = \boldsymbol{\Sigma}$ is a Gauss-Markov model if $Var(\mathbf{Y}) = \boldsymbol{\sigma}^2 I$.

(B) (8 points) For a linear model $E(\mathbf{Y}) = X\boldsymbol{\beta}$ with $Var(\mathbf{Y}) = \boldsymbol{\Sigma}$, a linear function of the parameters $\mathbf{c}^T\boldsymbol{\beta}$ is estimable if there is a vector of constants $\mathbf{a}$ such that $E(\mathbf{a}^T\mathbf{Y}) = X\hat{\beta}$, i.e., there is a linear unbiased estimator for $\mathbf{c}^T\boldsymbol{\beta}$.

(i) $\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3$ is not estimable. To obtain $\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3 = E(a_2 Y_{2j} + a_3 Y_{3k})$
$= a_2(\boldsymbol{\mu} + \boldsymbol{\alpha}_2) + a_3(\boldsymbol{\mu} + \boldsymbol{\alpha}_3) = (a_2 + a_3)\boldsymbol{\mu} + a_2\boldsymbol{\alpha}_2 + a_3\boldsymbol{\alpha}_3$ we must have $\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3 = 0$.

(ii) $2\boldsymbol{\mu} + 4\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_3$ is estimable because $2\boldsymbol{\mu} + 4\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_3 = E(4Y_{11} - Y_{21} - Y_{31})$.

Some students used the result that a linear function of the parameters $\mathbf{c}^T\boldsymbol{\beta}$ is estimable if and only if $\mathbf{c}^T\mathbf{d} = \mathbf{0}$ for <u>every</u> $\mathbf{d}$ such that $X\mathbf{d} = \mathbf{0}$. This can be used to show that $\alpha_2 + \alpha_3 = (0\ 0\ 1\ 1\ 0)\hat{\beta}$ is not estimable in part (i) by noting that $X\mathbf{d} = \mathbf{0}$ for $\mathbf{d} = (1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1)^T$ and $(0\ 0\ 1\ 1\ 0)\mathbf{d} = -2 \neq 0$. To use this result in part (ii) you must first identify all $\mathbf{d}$ vectors such that $X\mathbf{d} = \mathbf{0}$. In this case, the model matrix, call it X, has five columns and rank=4, so any $\mathbf{d}$ such that $X\mathbf{d} = \mathbf{0}$ must be of the form $\mathbf{d} = \omega(1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1)^T$ for some scalar ω. If there were another linearly independent vector, say $\mathbf{u}$, such that $X\mathbf{u} = \mathbf{0}$, then the model matrix would have rank less than 4.

(C) (5 points) Since the first row is the sum of the last five rows and the lower left corner is a 5x5 diagonal matrix, if follows that

$$W^TW = \begin{bmatrix} 10 & 4 & 2 & 2 & 2 \\ 4 & 4 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 2 & 0 \\ 2 & 0 & 0 & 0 & 2 \end{bmatrix}$$

has rank 5. A generalized inverse is obtained by inverting the 5x5 diagonal matrix in the lower
left
corner and replacing the first row and first column with zeros, i.e.,

$$(W^TW)^- = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & .25 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & .5 & 0 \\ 0 & 0 & 0 & 0 & .5 \end{bmatrix}.$$

Then, a solution to the normal equations is

$$\hat{\boldsymbol{\alpha}} = (W^T W)^- W^T \mathbf{Y} = \begin{bmatrix} 0 \\ \overline{Y}_{1.} \\ \overline{Y}_{2.} \\ \overline{Y}_{3.} \\ \overline{Y}_{4.} \end{bmatrix}.$$

(D) (5 points) The quantity $3\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_3 - \boldsymbol{\alpha}_4$ is estimable because $3\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_3 - \boldsymbol{\alpha}_4 = E(3\overline{Y}_{1.} - \overline{Y}_{2.} - \overline{Y}_{3.} - \overline{Y}_{4.})$. The least squares estimator is $3\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_2 - \hat{\boldsymbol{\alpha}}_3 - \hat{\boldsymbol{\alpha}}_4 = 3\overline{Y}_{1.} - \overline{Y}_{2.} - \overline{Y}_{3.} - \overline{Y}_{4.}$.

(E) (5 points) For a Gauss-Markov model with $E(\mathbf{Y}) = X\boldsymbol{\beta}$ and $Var(\mathbf{Y}) = \boldsymbol{\sigma}^2 I$, the ordinary least squares estimator for an <u>estimable</u> quantity $\mathbf{c}^T\boldsymbol{\beta}$ is the unique best linear unbiased estimator for $\mathbf{c}^T\boldsymbol{\beta}$. This implies that $3\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_2 - \hat{\boldsymbol{\alpha}}_3 - \hat{\boldsymbol{\alpha}}_4 = 3\overline{Y}_{1.} - \overline{Y}_{2.} - \overline{Y}_{3.} - \overline{Y}_{4.}$ is the unique best linear unbiased estimator for $3\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_3 - \boldsymbol{\alpha}_4$. The variance of any other unbiased estimator that is a linear function of $\mathbf{Y}$ is larger than $\sigma^2(2.25 + .5 + .5 + .5) = 3.75\sigma^2$.

(F) (i) (8 points) We can show that $\frac{1}{\boldsymbol{\sigma}^2} SSE_{model\,A} = \frac{1}{\boldsymbol{\sigma}^2} Y^T(I - P_X)Y$ has a non-central chi-squared distribution by showing that the conditions of Result 1 are satisfied. In this case $A = \frac{1}{\boldsymbol{\sigma}^2}(I - P_X)$ is symmetric and $\boldsymbol{\Sigma} = Var(\mathbf{Y}) = \boldsymbol{\sigma}^2 I$. Then, $A\boldsymbol{\Sigma} = (I - P_X)$ is clearly idempotent. Furthermore, the degrees of freedom are $rank(A) = rank(I - P_X) = n - rank(X) = 10 - 2 = 8$. The non-centrality parameter is $\boldsymbol{\delta}^2 = \frac{1}{\boldsymbol{\sigma}^2}\boldsymbol{\beta}^T W^T (I - P_X) W\boldsymbol{\beta}$ which is not necessarily zero because some columns of W are not linear combinations of the columns of X.

(ii) (8 points) Show that the conditions of Result 1 are satisfied with $A = \frac{1}{\boldsymbol{\sigma}^2}(I - P_X)P_W(I - P_X)$ and $\boldsymbol{\Sigma} = Var(\mathbf{Y}) = \boldsymbol{\sigma}^2 I$. Clearly, A is symmetric. To show that $A\boldsymbol{\Sigma} = (I - P_X)P_W(I - P_X)$ is idempotent, note that since each column of X is a linear combination of the columns of W then $P_W X$ and $P_W P_X = P_X = P_X P_W$. Consequently, $A\boldsymbol{\Sigma} = (I - P_X)P_W(I - P_X) = (I - P_X)(P_W - P_W P_X) = (I - P_X)(P_W - P_X) = (I - P_X)P_W = P_W - P_X$, which is easily shown to be an idempotent matrix. The degrees of freedom are $rank(A) = rank(P_W - P_X) = rank(W) - rank(X) = 4 - 2 = 2$. The noncentrality parameter, $\boldsymbol{\delta}^2 = \frac{1}{\boldsymbol{\sigma}^2}\boldsymbol{\beta}^T W^T (I - P_X) P_W (I - P_X) W\boldsymbol{\beta} = \frac{1}{\boldsymbol{\sigma}^2}\boldsymbol{\beta}^T W^T (P_W - P_X) W\boldsymbol{\beta}$, is not necessarily zero because some columns of W are not linear combinations of the columns of X and $P_X W$ is not W.

(G) (10 points) From part F(ii) we have that $\frac{1}{\boldsymbol{\sigma}^2}\mathbf{Y}^T(I - P_X)P_W(I - P_X)\mathbf{Y} = \frac{1}{\boldsymbol{\sigma}^2}\mathbf{Y}^T(P_W - P_X)\mathbf{Y}$ has a noncentral chi-squared distribution with 2 degrees of freedom. Use Result 1 to show that $\frac{1}{\boldsymbol{\sigma}^2} SSE_{model\,B} = \frac{1}{\boldsymbol{\sigma}^2}\mathbf{Y}^T(I - P_W)\mathbf{Y}$ has a central chi-squared distribution with 5 degrees of freedom. It follows from Result 2 that these two quadratic from are independent because

$$\begin{aligned}(I - P_X)P_W(I - P_X)(\sigma^2 I)(I - P_W) &= \sigma^2(I - P_X)P_W(I - P_X)(I - P_W)\\ &= \sigma^2(I - P_X)P_W(I - P_W - P_X + P_X P_W)\\ &= \sigma^2(I - P_X)P_W(I - P_W - P_X + P_X)\\ &= \sigma^2(I - P_X)P_W(I - P_W)\\ &= \sigma^2(I - P_X)(P_W - P_W)\end{aligned}$$

is a matrix of zeros. It follows that the F statistic has a noncentral F-distribution with (2,6) degrees of freedom if we divide the quadratic forms by their respective degrees of freedom. Consequently, c=6/2=3. The noncentrality parameter is shown in the solution to part F(ii).

(H) (5 points) Yes, it provides a goodness-of-fit test for model A. The null hypothesis is that mean milk production changes along a straight line as the level of the amino acid increases. The alternative is that mean milk production changes in some other way as the level of the amino acid increases.

Comment: A number of students incorrectly claimed that Model B is a reparameterization of Model A and incorrectly stated that $P_X = P_W$. Since rank(W)=4 and rank(X)=2, not every column of W can be written as a linear combination of the columns of X, and Model B is not a reparameterization of Model A. It is true that each column of X is a linear combination of the columns of W. Consequently, $P_W X = X$ and $P_X = P_W P_X = P_X P_W$.

2. (A) (5 points) $\mathbf{g}_2$ is estimable for model C if at least two of the values for Z_{11}, Z_{12}, Z_{13}, or Z_{14} are different, i.e., if any two cows in the control group are of different ages.

(B) (8 points) Write the null hypothesis as $H_0 : C\tilde{\mathbf{a}} = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}\begin{bmatrix}\gamma_0\\ \gamma_1\\ \gamma_2\\ \gamma_3\\ \gamma_4\\ \gamma_5\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$, and let

$\mathbf{b} = (M^T M)^{-1} M^T \mathbf{Y}$ be the solution to the normal equations. The null hypothesis would be rejected if

$$F = \frac{\dfrac{\mathbf{b}^T C^T (C^T (M^T M)^{-1} C)^{-1} C\mathbf{b}}{3}}{\dfrac{\mathbf{Y}^T (I - P_M)\mathbf{Y}}{4}}$$

exceeds the upper α percentile of a central F-distribution with (3, 4) degrees of freedom, where α is the type I error level of the test.

(C) (8 points) Compute the residual sum of squares for model C, i.e., $SSE_{\text{model C}} = \mathbf{Y}^T(I - P_M)\mathbf{Y}$,

where $P_M = M(M^T M)^- M^T$. If any two of the values for Z_{11}, Z_{12}, Z_{13}, or Z_{14} are different and

$Z_{21} \neq Z_{22}$ and $Z_{31} \neq Z_{32}$ and $Z_{41} \neq Z_{42}$, then $\text{rank}(I - P_M) = N - \text{rank}(M) = 10 - 6 = 4$ and from Result 1 we have that $SSE_{\text{model C}} = \mathbf{Y}^T(I - P_M)\mathbf{Y}$ has a central chi-square distribution with 4 degrees of freedom. Then,

$$0.95 = \Pr\{\chi^2_{4,.975} \leq \frac{SSE_{\text{model C}}}{\sigma^2} \leq \chi^2_{4,.025}\} \quad \Rightarrow \quad 0.95 = \Pr\left\{\frac{1}{\chi^2_{4,.025}} \leq \frac{\sigma^2}{SSE_{\text{model C}}} \leq \frac{1}{\chi^2_{4,.975}}\right\}$$

$$\Rightarrow \quad 0.95 = \Pr\left\{\frac{SSE_{\text{model C}}}{\chi^2_{4,.025}} \leq \sigma^2 \leq \frac{SSE_{\text{model C}}}{\chi^2_{4,.975}}\right\}$$

and a 95% confidence interval for σ^2 is $\left\{\frac{SSE_{\text{model C}}}{\chi^2_{4,.025}}, \frac{SSE_{\text{model C}}}{\chi^2_{4,.975}}\right\}$.

(D) (5 points) Model C is not a reparameterization of model B if any two of the values for Z_{11}, Z_{12}, Z_{13}, or Z_{14} are different, or $Z_{21} \neq Z_{22}$, or $Z_{31} \neq Z_{32}$, or $Z_{41} \neq Z_{42}$. For example, the last column of M is a linear combination of the columns of W if and only if there are constants a and b such that $a + b = Z_{41}$ and $a + b = Z_{42}$. This is impossible when $Z_{41} \neq Z_{42}$. The columns of W allow you to fit a different mean for milk production for each level of amino acid, but they provide no information on the age of the cows. Consequently, W does not provide enough information to allow you to fit trends in mean milk production across ages of cows within levels of amino acid.

3. (A) (5 points) $E(\mathbf{c}^T\mathbf{b}) = E(\mathbf{c}^T(X^TX)^{-1}X^T\mathbf{Y}) = \mathbf{c}^T(X^TX)^{-1}X^TE(\mathbf{Y}) = \mathbf{c}^T(X^TX)^{-1}X^TX\boldsymbol{\beta} = \mathbf{c}^T\boldsymbol{\beta}$.

(B) (10 points) Since $Var(\mathbf{Y}) = \sigma^2 I + \delta^2\mathbf{1}\mathbf{1}^T$ does not satisfy the Gauss-Markov property, it is not obvious that the OLS estimator=r, $\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^{-1}X^T\mathbf{Y})$, is a best linear unbiased estimator for $\mathbf{c}^T\boldsymbol{\beta}$. As shown in the lecture notes, the generalized least squares estimator,

$$\mathbf{c}^T\mathbf{b}_{GLS} = \mathbf{c}^T(X^T(\sigma^2 I + \delta^2\mathbf{1}\mathbf{1}^T)^{-1}X)^{-1}X^T(\sigma^2 I + \delta^2\mathbf{1}\mathbf{1}^T)^{-1}\mathbf{Y})$$

is the unique best linear unbiased estimator for $\mathbf{c}^T\boldsymbol{\beta}$. Then, $\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^{-1}X^T\mathbf{Y})$ is also a best linear unbiased estimator if and only if it is equal to $\mathbf{c}^T\mathbf{b}_{GLS}$. Intuitively this seems possible, because the variances are equal and all covariances are equal. Consequently, the GLS estimator will weight the observations in the same manner as the OLS estimator. One way to show this is to recall that $(\sigma^2 I + \delta^2\mathbf{1}\mathbf{1}^T)^{-1} = \frac{1}{\sigma^2}(I - \frac{\delta^2}{\sigma^2 + n\delta^2}\mathbf{1}\mathbf{1}^T)$ and insert this into the formula for $\mathbf{c}^T\mathbf{b}_{GLS}$ to show that it reduces to $\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^{-1}X^T\mathbf{Y})$. I did not anticipate that anyone would do this. An alternative approach is to directly show that OLS estimator is the linear unbiased estimator with the smallest variance. Consider any other linear unbiased estimator, say $\mathbf{d}^T\mathbf{Y}$, for $\mathbf{c}^T\boldsymbol{\beta}$. Then, $\mathbf{c}^T\boldsymbol{\beta} = E(\mathbf{d}^T\mathbf{Y}) = \mathbf{d}^TX\boldsymbol{\beta}$ implies that $\mathbf{c}^T = \mathbf{d}^TX$. Furthermore,

$$Var(\mathbf{d}^T\mathbf{Y}) = \mathbf{d}^T(\sigma^2 I + \delta^2\mathbf{1}\mathbf{1}^T)\mathbf{d} = \sigma^2\mathbf{d}^T\mathbf{d} + \delta^2\mathbf{d}^T\mathbf{1}\mathbf{1}^T\mathbf{d}.$$

It is easy to see that the variance of the OLS estimator is

$$
\begin{aligned}
\mathrm{Var}(\mathbf{c}^T\mathbf{b}) &= \mathbf{c}^T(X^TX)^{-1}X^T(\sigma^2 I + \delta^2\mathbf{1}\mathbf{1}^T)X(X^TX)^{-1}\mathbf{c} \\
&= \sigma^2\mathbf{c}^T(X^TX)^{-1}\mathbf{c} + \delta^2\mathbf{c}^T(X^TX)^{-1}X^T\mathbf{1}\mathbf{1}^T X(X^TX)^{-1}\mathbf{c} \\
&= \sigma^2\mathbf{d}^T X(X^TX)^{-1}X^T\mathbf{d} + \delta^2\mathbf{d}^T X(X^TX)^{-1}X^T\mathbf{1}\mathbf{1}^T X(X^TX)^{-1}X^T\mathbf{d} \qquad \text{since } \mathbf{c}^T = \mathbf{d}^T X \\
&= \sigma^2\mathbf{d}^T P_X\mathbf{d} + \delta^2\mathbf{d}^T\mathbf{1}\mathbf{1}^T\mathbf{d} \qquad \text{since } P_X\mathbf{1} = \mathbf{1} \\
&= \sigma^2(P_X\mathbf{d})^T P_X\mathbf{d} + \delta^2\mathbf{d}^T\mathbf{1}\mathbf{1}^T\mathbf{d} \qquad \text{since } P_X P_X = P_X \text{ and } P_X^T = P_X \\
&\le \sigma^2\mathbf{d}^T\mathbf{d} + \delta^2\mathbf{d}^T\mathbf{1}\mathbf{1}^T\mathbf{d} = \mathrm{Var}(\mathbf{d}^T\mathbf{Y}) \qquad \text{since } (P_X\mathbf{d})^T P_X\mathbf{d} \le \mathbf{d}^T\mathbf{d}\text{,}
\end{aligned}
$$

the length of the projection of $\mathbf{d}$ onto the space spanned by the columns of X is no longer than the length of $\mathbf{d}$.

Consequently, the OLS estimator is a minimum variance linear unbiased estimator for $\mathbf{c}^T\hat{\mathbf{a}}$ in this case, even though the Gauss-Markov property is not satisfied.

EXAM SCORES:

```
90 | 5
90 |
80 | 5 6 6 6 7
80 | 0 3 4 4 4
70 | 5 6 6 6 7 7 8 9
70 | 0 0 1 3 3 3 3 3 4 4 4
60 | 5 5 5 5 5 6 6 6 6 7 7 8 8 9 9
60 | 0 0 1 3 4 4
50 | 5 5 6 6 7 7 9 9 9
50 | 0 0 1 2 4 4
40 | 6 9
40 |
```

A point value should be shown for the credit awarded for each part of your exam, corresponding to the point values listed above. If your answer failed to earn any credit, a zero should be shown. If no point value is shown for some part of your exam, show your exam to the instructor. Also, check if the point total recorded on the last page of your exam is correct.