In the analysis of linear models with missing data, SAS GLM TYPE III and TYPE IV sums of squares are different for factor comparisons which contain the missing cells. To determine the hypotheses being tested for each type of sums of squares, GLM will print the estimable functions for the effects of the beta or over parameterized model; however, these are difficult to interpret. The cell means (full rank) model is presented as a tool for evaluating the hypotheses being tested in terms of linear combinations of population cell means. These contrasts are easy to interpret but do not suggest which type of sums of squares should be chosen. The choice of which set of sums of squares are appropriate is a task left to the researcher and should be based on the questions and requirements of the experiment. An experiment from the Corps of Engineers Dredged Materials project is given as an example.

INTRODUCTION

The analysis of linear models with missing data has always been a dilemma for statisticians. Recently a number of statistical software packages have been developed to accommodate this situation (see Speed et al. 1978 and Freund 1980). Each program utilizes a different algorithm which usually vary in the assumptions made concerning the missing data. The widely used General Linear Model (GLM) procedure of the Statistical Analysis System (SAS) exemplifies this problem by offering the user four different ANOV tables denoted by the four different types of sums of squares. The question then arises, which type of sums of squares, if any, should be used? There is no immediate answer to this question because the purpose and nature of the experiment to a large degree dictates which hypotheses are "appropriate."

To aid the user in making a choice, SAS GLM will print the estimable functions for each type of sums and squares. The hypothesis matrix can then be generated yielding the linear combinations of effects (from the original model) which are contained in each type of sums of squares.

The purpose of this article is to present an alternate method for evaluating the hypotheses being tested by SAS GLM TYPE III and IV sums of squares. It is suggested that by translating the tests of hypotheses into linear combinations of cell means, the user of GLM is better able to understand the specific hypotheses being made. This approach incorporates the statistical equivalence between two linear models—the beta and the cell means model.

THE MODELS

BETA

Consider the two-way classification fixed effects model with interaction. The linear additive model would be:

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + \epsilon_{ijk} \]  

where

- \( Y_{ijk} \) = value of \( k \)th observation taken on some random variable from the \( j \)th level of \( B \) and the \( i \)th level of \( A \)
- \( \mu \) = overall mean
- \( \alpha_i \) = effect of \( i \)th level of factor \( A \) with \( i = 1, 2, \ldots, a \)
- \( \beta_j \) = effect of \( j \)th level of factor \( B \) with \( j = 1, 2, \ldots, b \)
- \( (\alpha \beta)_{ij} \) = interaction of \( ij \)th combination of \( A \) and \( B \)
- \( \epsilon_{ijk} \) = random error assumed NID \((0, \sigma^2)\).

In matrix notation the above model is represented by:

\[ Y = XB + \epsilon \]  

where

- \( Y \) is an \( nx1 \) vector of observations
- \( X \) is an \( nx(l+a+b+ab) \) incidence matrix of ones and zeros
- \( B \) is an \( (l+a+b+ab) \times 1 \) vector of unknown parameters
- \( \epsilon \) is an \( nx1 \) vector of random errors assumed to be distributed \((0, \sigma^2)\).

This linear model shall be referred to as the "Beta" or the over parameterized model. The least squares normal equations are given by:

\[ (X'X)b = (X'Y) \]  

where

- \( b \) is the \((l+a+b+ab) \times 1 \) solution vector
- \( X'X \) is an \((l+a+b+ab) \times (l+a+b+ab) \) incidence matrix
- \( X'Y \) is an \((l+a+b+ab) \times 1 \) vector of cross-products.

However, the \((X'X)\) incidence (design) matrix is singular of rank \( q < r \). In this over parameterized model, the number of independent equations of \((X'X)\) is less than the number of unknown parameters of the solution vector; therefore, \((X'X)^{-1}\) does not exist. The solution must then be obtained by an alternate method. Two such methods are 1) the imposition of restrictions on the model parameters, and 2) the use of a generalized inverse or no restriction method.

The SAS GLM procedure operates on the beta model and allows the user to specify the effects (factors) which are to be contained within the model. To solve the least squares normal equations, PROC GLM employs a \( g2 \) generalized inverse

547
This type of generalized inverse is obtained by setting the last effect in each partition of the \((X'X)\) matrix to zero and then inverting the resulting matrix. Unfortunately, this solution enables only certain linear combinations of the solution vector elements to be estimated or contrasted. Thus the concept and confusion regarding estimability arises. A linear combination of the solution vector, \(q'b\), is estimable if and only if a linear combination of the expected value of the observation vector, \(\bar{t}'E(Y)\), is equal to \(q'b\) (Graybill 1961, Searle 1971, Goodnight 1976). The set of such linear combinations for GLM computations is given by the general form of the estimable functions as determined by the product:

\[(X'X)^{-1}(X'X)\]  

where \((X'X)^{-1}\) is the \(g^2\) generalized inverse used by SAS GLM. Therefore, \(q'b\) is estimable if a generating set of linear combinations \((L's)\) from \((X'X)(X'X)\) are estimable. The general form of the estimable functions and TYPE III and TYPE IV estimable functions may be obtained by specifying the \(E, E3,\) and \(E4\) options in the model statement of GLM.

Since the matrix of estimable functions is of full row rank, the number of symbols (the \(L's\) of the estimable functions) represent the maximum rank of any hypothesis matrix generated. The original \(L\) matrix being tested by each type of sums of squares may be reconstructed by setting each "free" coefficient of the estimable functions, in turn, to a value of one and then zero (Searle 1971, Goodnight 1980). Hence, by obtaining the estimable functions the hypothesis matrix \((L)\) in terms of model effects may be determined.

In the case of no missing cells, TYPE III and TYPE IV estimable functions are identical. That is, they are testing the same set of hypotheses. However, if missing cells are present, the estimable functions for the interaction terms of \(E, E3,\) and \(E4\) options in the model statement of GLM may be obtained by specifying the \(E, E3,\) and \(E4\) options in the model statement of GLM.

An equivalent form of the two-way classification fixed effects model as expressed by the cell means model would be:

\[Y_{ijk} = \mu_{ij} + \epsilon_{ijk}\]  

where \(\mu_{ij}\) is the cell mean of the \(ij\)th combination of \(AB\)

In matrix notation the cell means model may be written as:

\[Y = \mathbf{W}_i + \epsilon\]  

where \(Y\) is an \(nx1\) vector of observations taken on some random variable

\(W\) is an \(nx(a\times b)\) incidence matrix of ones and zeros denoting from which population the sample was taken.

\(\mu\) is an \((a\times b)\times 1\) vector of unknown parameters.

\(\epsilon\) is an \(nx1\) vector of unknown parameters.

\(G\) is a \(rx(a\times b)\) constraint matrix displaying some assumed or known relationship between the population parameters.

In the cell means model the assumption is made that a sample is taken from \((a\times b)\) univariate populations, each with a mean and variance (Speed 1969, Hocking and Speed 1975). The elements of the least squares solution vector are estimates of these population or cell means. The least squares solution of the normal equations is direct. The \(W'W\) matrix is diagonal and of full rank with the elements being the number of times the \(ij\)th population \((i.e., \mu_{ij})\) was sampled, therefore, the \((W'W)^{-1}\) matrix exists. The cell means model thus avoids the imposition of non-estimable conditions on the population parameters as required by the beta model. In addition, any linear combination of the solution vector elements is estimable as long as the population has been sampled (Speed 1969).

The constraint matrix, \(G\), is a hypothesis matrix testing some known or assumed relationship between the population parameters. For example, if in the two-way classification model it is assumed that there is no \((a\times b)\) interaction, the constraint matrix would be a set of multipliers to test for no \((a\times b)\) interaction, or,

\[\mu_{ij} = \mu_{i1} + \mu_{1j} + \mu_{11} = 0\]  

for all \(ij\)

**COMPARISON BETWEEN THE BETA AND CELL MEANS MODELS**

The beta and cell means models as described above are statistically equivalent tests, that is, there is a one-to-one correspondence between testable hypotheses and linear estimable functions of the parameters (Speed 1969). The two models are related by the function

\[\mu = \mathbf{PS}\]  

where
and $S$ are the solution vectors of the cell means and beta models respectively.

$P$ is the design matrix for the effects in the beta model.

For example, for a two-way factorial experiment with two levels of factor $A$ and of factor $B$, with no interaction, the two models are then related by

$$
\begin{align*}
\mu & = P \alpha + S \beta \\
\mu_{ij} & = 1 1 0 1 0 \mu \\
\mu_{i2} & = 1 1 0 0 1 \alpha_1 \\
\mu_{2j} & = 1 0 1 1 0 \alpha_2 \\
\mu_{ij2} & = 1 0 1 0 1 \beta_1 \\
\mu_{i2j} & = 1 0 1 0 1 \beta_2
\end{align*}
$$

Since $P$ is the design matrix, the effects of the beta model which are to be included must be stated explicitly.

If $H\mu = 0$, where $H$ is the hypothesis matrix, is being tested in the cell means model, the specific linear combinations of population parameters which are being tested in a beta model for the same set of data are given by

$$
\delta\beta = 0 \quad (9)
$$

Since $\mu = P\alpha$, $H\mu = 0$ may be written as $HP\alpha = 0$.

Letting HP be denoted by $S$, equation (9) is obtained.

To interpret the hypotheses being tested by TYPE III and TYPE IV SS in terms of the cell means given $S\beta = 0$, the hypothesis matrix for $H\mu = 0$ is determined by

$$
H = SP^+ \quad (10)
$$

where

- $P^+$ is the generalized inverse of the design matrix $P$
- $S$ is the coefficient matrix of the estimable functions obtained from SAS GLM.

This relationship follows since $HP = 0$. Therefore, the hypotheses being tested by each type of sums of squares can be translated from the effects of the beta model to linear combinations of the cell means. When going from the beta model to the cell means model there are two cases to consider.

Case 1 is the unconstrained model or when no assumptions are made concerning any parameters within the model. The hypothesis matrix for the cell means model is then equal to the coefficient matrix of the estimable functions for the highest order interaction. This is the simplest case with no additional computations required. The hypothesis matrix can be constructed directly by determining the original $L$ matrix for the highest order interaction. Case 2, the constrained model, assumes that an interaction term or group of effects are equal to zero or some constant. For this situation, the hypothesis matrix must be generated using equation (10). This can be easily accomplished using the PROC MATRIX program in SAS.

**AN EXAMPLE**

The concepts developed in the preceding discussion were applied to a data base obtained from the Corps of Engineers Waterways Experiment Station in Vicksburg, Mississippi. The analysis presented here is based on a subset of data from the Galveston Bay, Texas Dredged Materials project. This study was designed to investigate the possibility of recolonizing dredged bank material with two species of cord grass, Spartina alterniflora and Spartina patens. The experimental design is a 2x2x3 factorial with missing cells. The factors of interest were species type, two levels of vertical elevation, and three levels of fertilizer. Due to logistical problems, the cells 212 and 222 were not sampled (Table 1).

<table>
<thead>
<tr>
<th>A</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>W111</td>
<td>W112</td>
</tr>
<tr>
<td></td>
<td>W211</td>
<td>W212</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1. Layout of experimental design showing factor levels, location of missing cells, and identification of cell means.

Factor A = species type  
Factor B = elevation type  
Factor C = level of fertilizer.

**CASE 1—THE UNCONSTRAINED MODEL**

The unconstrained beta model is written as

$$
Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk} \quad (11)
$$

and the cell means model as

$$
Y_{ijk} = \mu + \epsilon_{ijk} \quad (12)
$$

The SAS GLM output for this example is given in Table 2.

*Note: $X$ = missing cells.*
Table 2. SAS GLM output with TYPE III and TYPE IV SS for unconstrained model

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Type IV SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>9</td>
<td>17241.89238165</td>
<td>1915.76538165</td>
<td>22.51</td>
</tr>
<tr>
<td>Error</td>
<td>28</td>
<td>1167.38433054</td>
<td>58.32681918</td>
<td>PR &gt; F</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>29</td>
<td>18409.73276519</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>

R-Square: C.V. 0.890564 7.361167 87.230119

H: Source Type III SS F Value PR > F

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>F Value</th>
<th>Type IV SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>1</td>
<td>1959.86418443</td>
<td>33.56 0.0001</td>
</tr>
<tr>
<td>Elev</td>
<td>2</td>
<td>1792.36919208</td>
<td>15.20 0.0001</td>
</tr>
<tr>
<td>FertType</td>
<td>2</td>
<td>1475.51987733</td>
<td>6.39 0.0046</td>
</tr>
<tr>
<td>SpeciesxElev</td>
<td>2</td>
<td>1792.61977719</td>
<td>23.68 0.0001</td>
</tr>
<tr>
<td>SpeciesxFertType</td>
<td>2</td>
<td>1879.38068148</td>
<td>1.61 0.2448</td>
</tr>
<tr>
<td>SpeciesxElevxFertType</td>
<td>1</td>
<td>10.66938118</td>
<td>0.10 0.7739</td>
</tr>
</tbody>
</table>

The estimable functions for the factor species, which are the same for TYPE III and TYPE IV SS, are:

**TYPE III AND IV ESTIMABLE FUNCTIONS FOR SPECIES**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficients</th>
<th>Intercepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Species</td>
<td>0 12</td>
<td></td>
</tr>
<tr>
<td>Elev</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>FertType</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>SpeciesxElev</td>
<td>0.5112 0.5112</td>
<td></td>
</tr>
<tr>
<td>SpeciesxFertType</td>
<td>-0.5112 -0.5112</td>
<td></td>
</tr>
<tr>
<td>ElevxFertType</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>SpeciesxElevxFertType</td>
<td>0.2512 0.2512</td>
<td></td>
</tr>
</tbody>
</table>

Therefore, the hypothesis matrix for the beta model in terms of factor levels or model effects is (setting L2 to 4):

\[ H = \begin{bmatrix} 0 & 4 & -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & -2 & -2 & 0 & 2 & 2 & -2 & -2 \end{bmatrix} \]

or

\[ 4a_1 + 2(a_{B11} + a_{B12} + a_{Y11} + a_{Y12}) + (a_{BY111} + a_{BY112} + a_{BY121} + a_{BY122}) + -4a_2 + 2(a_{B21} + a_{B22} + a_{Y21} + a_{Y22}) + (a_{BY211} + a_{BY212} + a_{BY221} + a_{BY222}) \]

\[ (13) \]

Since we are in the unconstrained model, the hypothesis matrix for the cell means model is constructed from the coefficient matrix of the estimable functions for the interaction species* elevation*fertilizer type (A*B*C). Hence, the hypothesis being tested by both TYPE III and TYPE IV SS for the factor species in terms of the cell means (setting L2 to 1) is:

\[ 0.25 (\mu_{1111} + \mu_{1112} + \mu_{1211} + \mu_{1212}) = 0.25 (\mu_{1111} + \mu_{1112} + \mu_{2111} + \mu_{2112}) \]

\[ (14) \]

or,

\[ \begin{bmatrix} \mu_{1111} & \mu_{1112} & \mu_{1121} & \mu_{1122} & \mu_{1211} & \mu_{1212} & \mu_{2111} & \mu_{2112} \end{bmatrix} = \begin{bmatrix} -\mu_{1111} & X & -\mu_{1112} & \mu_{1121} & \mu_{1211} & \mu_{1212} & -\mu_{2111} & X \end{bmatrix} \]

550
TYPE III and TYPE IV SS are different for the factor fertilizer. Again, they are testing a different set of hypotheses. The estimable functions differ with respect to the coefficients given to the factor interactions (Table 3).

In terms of the effects of the beta model TYPE III SS contain:

\[ y_1 = 0.5(a_{Y11} + a_{Y21} + a_{Y12} + a_{Y22}) + 0.25(a_{Y11} + a_{Y12} + a_{Y21} + a_{Y22}) \]

\[ y_2 = 0.5(a_{Y13} + a_{Y23} + a_{Y12} + a_{Y22}) + 0.25(a_{Y11} + a_{Y12} + a_{Y21} + a_{Y22}) \]

\[ y_2 = a_{Y12} + 0.25(a_{Y21}) + 0.5(a_{Y12} + a_{Y22} + a_{Y11} + a_{Y12}) + 0.125(a_{Y11} + a_{Y22}) = y_3 + 0.75(a_{Y13} + a_{Y23}) + 0.25(a_{Y11} + a_{Y22}) + 0.5(a_{Y12} + a_{Y22}) + 1.25(a_{Y11} + a_{Y12} + a_{Y21} + a_{Y22}) + 0.375(a_{Y11} + a_{Y12} + a_{Y21} + a_{Y22}) \]

The TYPE IV SS contain

\[ y_1 = 0.5(a_{Y11} + a_{Y21} + a_{Y12} + a_{Y22}) + 0.25(a_{Y11} + a_{Y12} + a_{Y21} + a_{Y22}) = y_3 + 0.5(a_{Y13} + a_{Y23} + a_{Y12} + a_{Y22}) + 0.25(a_{Y11} + a_{Y12} + a_{Y21} + a_{Y22}) \]

\[ y_1 = a_{Y12} + 0.5(a_{Y12} + a_{Y22} + a_{Y11} + a_{Y22}) = y_3 + a_{Y13} + 0.5(a_{Y12} + a_{Y22} + a_{Y11} + a_{Y22}) \]

This form of the sums of squares is difficult to interpret and even though the first row of the hypothesis matrices are the same, the benefit of choosing either TYPE III or TYPE IV SS is not apparent. However, if we translate these hypotheses into linear contrasts of the cell means (by using $H = P^T$), the exact nature of the tests is clear. In the cell means model TYPE III SS are testing

\[ H_1 + H_2 + H_1 - H_2 = H_1 + H_3 + H_2 - H_3 \]

or,

\[ H_1 \]

\[ H_2 \]

\[ H_3 \]

Table 3. Estimable functions for fertilizer

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>TYPE III COEFFICIENTS</th>
<th>TYPE IV COEFFICIENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SPECIES</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>ELEU</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FERTTYPE</td>
<td>L6</td>
<td>L6</td>
</tr>
<tr>
<td>SPECIES*ELEU</td>
<td>-L6-L7</td>
<td>-L6-L7</td>
</tr>
<tr>
<td>SPECIES+FERTTYPE</td>
<td>0.5L6-0.25L7</td>
<td>0.5L6-0.25L7</td>
</tr>
<tr>
<td></td>
<td>-0.5L6-0.25L7</td>
<td>-0.5L6-0.25L7</td>
</tr>
<tr>
<td>ELEU+FERTTYPE</td>
<td>0.5L6 0.5L7</td>
<td>0.5L6 0.5L7</td>
</tr>
<tr>
<td>SPECIES*ELEU+FERTYPE</td>
<td>0.5L6-0.5L7</td>
<td>0.5L6-0.5L7</td>
</tr>
<tr>
<td></td>
<td>0.25L6 0.5L6</td>
<td>0.25L6 0.5L6</td>
</tr>
<tr>
<td></td>
<td>-0.25L6 0.75L6</td>
<td>-0.25L6 0.75L6</td>
</tr>
<tr>
<td></td>
<td>0.5L6 0.5L7</td>
<td>0.5L6 0.5L7</td>
</tr>
<tr>
<td></td>
<td>-0.5L6-0.5L7</td>
<td>-0.5L6-0.5L7</td>
</tr>
</tbody>
</table>
and,

\[ 4(U_{112} + U_{121}) + (U_{211} + U_{221}) = (U_{111}U_{121}U_{123}U_{223}) + 3(U_{113}U_{123}) \]  

(18)

or,

\[
\begin{array}{ccc}
U_{111} & + & U_{121} & - & U_{123} \\
- & U_{112} & + & U_{122} & - & U_{123} \\
U_{211} & + & U_{221} & - & U_{223} \\
- & U_{212} & + & U_{222} & - & U_{223} \\
\end{array}
\]

\[
\begin{array}{ccc}
X & & X & - & U_{123} \\
& & X & - & U_{123} \\
& & X & - & U_{223} \\
& & X & - & U_{223} \\
\end{array}
\]

Relative to the cell means model TYPE IV SS are testing

\[ U_{111}U_{121}U_{123}U_{221} = U_{113} + U_{123} + U_{213} + U_{223} \]  

(19)

(which is equal to the first row of the hypothesis matrix being tested by TYPE III SS)

and,

\[ U_{113}U_{123} = U_{113} + U_{123} \]  

(20)

or,

\[
\begin{array}{ccc}
0 & + & U_{112} & - & U_{123} \\
0 & + & U_{122} & - & U_{123} \\
0 & + & U_{211} & - & U_{223} \\
0 & + & U_{222} & - & U_{223} \\
\end{array}
\]

\[
\begin{array}{ccc}
X & & X & - & U_{123} \\
X & & X & - & U_{223} \\
X & & X & - & U_{123} \\
X & & X & - & U_{223} \\
\end{array}
\]

CASE 2 - THE CONSTRAINED MODEL

If we assume that the species*elevation*fertilizer interaction is zero, the linear model for the beta model becomes

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk} \]  

(21)

and the cell means model is written as

\[ Y_{ijk} = \mu_{ijk} + \epsilon_{ijk} \]  

(22)

subject to \( G_{ijk}=0 \) (no A*B*C interaction).

which is similar to the unconstrained model except for the addition of the constraint matrix \( G \). The SAS GLM output for this model is given in Table 4. The same sums of squares are obtained for fertilizer but the 3rd order interaction is no longer present in the estimable functions (Table 5). The hypotheses matrices for the cell means model must be determined using the equation \( H = G \) (equation 10). Invoking PROC MATRIX and reading in the coefficient matrix of the estimable functions (Appendix A) the hypothesis matrices for TYPE III and TYPE IV SS are:

**TYPE III**

| \( W_{111} U_{112} U_{113} U_{121} U_{122} U_{123} U_{211} U_{212} U_{221} U_{222} \) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | -1 | 1 | 0 | -1 | 1 | -1 | 1 | -1 | -1 |

Table 4. SAS GLM output with TYPE III and TYPE IV SS for the constrained model.

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>SUM OF SQUARES</th>
<th>MEAN SQUARE</th>
<th>F VALUE</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL</td>
<td>8</td>
<td>17231.45009457</td>
<td>2153.06251198</td>
<td>36.36</td>
<td></td>
</tr>
<tr>
<td>ERROR</td>
<td>21</td>
<td>1175.49767472</td>
<td>56.11860784</td>
<td>0.0006</td>
<td></td>
</tr>
<tr>
<td>CORRECTED TOTAL</td>
<td>29</td>
<td>18406.94767472</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-SQUARE | 0.936985 | 0.5205 | 7.49123540 | 87.92278001 |

**SOURCE** | DF | TYPE III SS | F VALUE | PR > F |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SPECIES</td>
<td>1</td>
<td>1959.80410443</td>
<td>34.02</td>
<td>0.0001</td>
</tr>
<tr>
<td>ELEV</td>
<td>1</td>
<td>656.00687078</td>
<td>108.45</td>
<td>0.0001</td>
</tr>
<tr>
<td>FERTTYPE</td>
<td>3</td>
<td>1725.35835858</td>
<td>15.97</td>
<td>0.0001</td>
</tr>
<tr>
<td>SPECIES*ELEV</td>
<td>1</td>
<td>1673.61897778</td>
<td>29.82</td>
<td>0.0001</td>
</tr>
<tr>
<td>SPECIES*FERTTYPE</td>
<td>1</td>
<td>147.58188777</td>
<td>2.62</td>
<td>0.1198</td>
</tr>
<tr>
<td>ELEV*FERTTYPE</td>
<td>3</td>
<td>187.98882314</td>
<td>1.67</td>
<td>0.2114</td>
</tr>
</tbody>
</table>

**SOURCE** | DF | TYPE IV SS | F VALUE | PR > F |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SPECIES</td>
<td>1</td>
<td>1959.80410443</td>
<td>34.02</td>
<td>0.0001</td>
</tr>
<tr>
<td>ELEV</td>
<td>1</td>
<td>656.00687078</td>
<td>108.45</td>
<td>0.0001</td>
</tr>
<tr>
<td>FERTTYPE</td>
<td>2</td>
<td>1899.69784296</td>
<td>10.78</td>
<td>0.0096</td>
</tr>
<tr>
<td>SPECIES*ELEV</td>
<td>1</td>
<td>1673.61897778</td>
<td>29.82</td>
<td>0.0001</td>
</tr>
<tr>
<td>SPECIES*FERTTYPE</td>
<td>1</td>
<td>147.58188777</td>
<td>2.62</td>
<td>0.1198</td>
</tr>
<tr>
<td>ELEV*FERTTYPE</td>
<td>2</td>
<td>187.98882314</td>
<td>1.67</td>
<td>0.2114</td>
</tr>
</tbody>
</table>

*Note: Other type Iu testable hypotheses exist which may yield different SS.
These hypothesis matrices yield linear combinations of the cell means which are identical to those obtained for the unconstrained model.

CONCLUSIONS AND RECOMMENDATIONS

By utilizing the estimable functions generated by the SAS GLM procedure, the cell means model can be a valuable aid in interpreting the hypotheses being tested by TYPE III and TYPE IV sums of squares. This procedure presents the contrasts in terms of linear combinations of the cell means, a form which is easier to interpret than the effects given by the beta model. The final decision as to which type of sums of squares are "best" must be left to the researcher. The method given here does not imply which sums of squares are appropriate, but rather aids the statistician in determining what SAS GLM procedure is actually testing.

Basic recommendations are related to the initial use of the cell means model. In this way, the user:

1) avoids the confusion regarding estimability and estimable functions.
2) need not work in the over parameterized model and be concerned with which hypotheses are being tested.
3) can obtain a solution vector which has an expected value of the population parameters, or cell means (see Koonce and Speed 1980).

The user may also avoid having to choose a set of sums of squares by using the CONTRAST and ESTIMATE statements in GLM. These options allow the user to specify directly the hypothesis matrix for testing or estimating linear combinations of the population parameters.

REFERENCES


APPENDIX A

Sample program for deriving the cell means model hypothesis matrix, constrained model, using the coefficient matrix of the SAS GLM estimable functions. Species, land elevation and amount of fertilizer are denoted by A, B, and C respectively. Data set old is the primary data base containing the input data.

DATA ONE; SET OLD;
   A=B=(A*10)+B;
   AC=(A*10)+C;
   BC=(B*10)+C;
DATA ALL; KEEP A B C AB AC BC;
PROC MATRIX;
FETCH X DATA=ALL;
   A = X(1,); B = X(2,); C = X(3,);
   AB = X(4,); AC = X(5,); BC = X(6,);
NOTE CONSTRUCTION OF DESIGN MATRIX F;
   F = J(11,1,1)||DESIGN(A)||DESIGN(B)||DESIGN(C)||
      DESIGN(AB)||DESIGN(AC)||DESIGN(BC);
NOTE GENERALIZED INVERSE OF F, F+;
   PG2 = GINV(P);
NOTE TYPE IV ESTIMABLE FUNCTIONS COEFFICIENT MATRIX;
   DELTA_A = 1 0 -1/ 0 1 -1;
   DELTA_B = 0.5 0 -0.5 0.5 0 -0.5 0.5 0
      -0.5/ 0.1 -1 0 0 0 1 -0.5 0 0.5 -0.5;
   DELTA = J(2,5,0)||DELTA_A||DELTA_B;
NOTE CELL MEANS HYPOTHESIS MATRIX-H;
   H = DELTA*PG2; PRINT H;