Appendix 11: Sums of Squares

This appendix describes methods for computing sums of squares.

Notation

The notation used in this appendix is the same as that in the GLM Univariate and Multivariate chapter.

Type I Sum of Squares and Hypothesis Matrix

The Type I sum of squares is computed by fitting the model in steps according to the order of the effects specified in the design and recording the difference in error sum of squares (ESS) at each step.

By applying the SWEEP operator on the rows and columns of the augmented matrix $\mathbf{Z'WZ}$ (of dimension $(p+r)\times(p+r)$), the Type I sum of squares and its hypothesis matrix for each effect (except for the intercept effect, if any) is obtained.

Calculating the Sum of Squares

The following procedure is used to find the Type I sum of squares for effect F:

Let the order of effects specified in the design be F_0, F_1, F_2, ..., F_m. The columns of $\mathbf{X}$ are partitioned into $\mathbf{X}_0$, $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_m$, where $\mathbf{X}_0 = \mathbf{1}$ corresponds to the intercept effect F_0, and the columns in the submatrix $\mathbf{X}_j$ correspond to effect F_j, $j = 0,1,\ldots,m$.

Let F_j be the effect F of interest. Let $\text{ESS}_{j-1}(l)$ and $\text{ESS}_j(l)$ be the lth diagonal elements of the $r\times r$ lower diagonal submatrix of $\mathbf{Z'WZ}$ after the SWEEP operator is applied to the columns associated with $\mathbf{X}_0$, $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_j$, $j = 0,1,\ldots,m$. When the lth column of $\mathbf{Y}$ is used as the dependent variable, the Type I sum of squares for effect F_j is $\text{ESS}_{j-1}(l) - \text{ESS}_j(l)$, $l = 1,\ldots,r$, where $\text{ESS}_{-1}(l)$ is defined as 0.

Constructing a Hypothesis Matrix

The hypothesis matrix **L** is constructed using the following steps:

1. Let $\mathbf{L}_0$ be the upper diagonal $p \times p$ submatrix of $\mathbf{Z}'\mathbf{WZ}$ after the SWEEP operator is applied to the columns associated with the effects *preceding* F. Set the *columns* and *rows* of $\mathbf{L}_0$, which are associated with the effects *preceding* F, to 0.
2. For the *rows* of $\mathbf{L}_0$ associated with the effects *ordered after* F_j, if any, set the corresponding rows of $\mathbf{L}_0$ to 0. Remove all of the 0 rows in the matrix $\mathbf{L}_0$. The row dimension of $\mathbf{L}_0$ is then less than *p*.
3. Use row operations on the rows of $\mathbf{L}_0$ to remove any linearly dependent rows. The set of all nonzero rows of $\mathbf{L}_0$ forms a Type I hypothesis matrix **L**.

Type II Sum of Squares and Hypothesis Matrix

A Type II sum of squares is the reduction in ESS due to adding an effect after all other terms have been added to the model except effects that *contain* the effect being tested.

For any two effects F and F', F is *contained* in F' if the following conditions are true:

- Both effects F and F' involve the same covariate, if any.
- F' consists of more factors than F.
- All factors in F also appear in F'.

Intercept Effect. The intercept effect μ is contained in all the pure factor effects. However, it is not contained in any effect involving a covariate. No other effect is contained in the intercept effect.

Calculating the Sum of Squares

To find the Type II (and also Type III and IV) sum of squares associated with any effect F, you must distinguish which effects in the model contain F and which do not. The columns of **X** can then be partitioned into three groups: $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$, where:

- $\mathbf{X}_1$ consists of columns of **X** that are associated with effects that do not contain F.
- $\mathbf{X}_2$ consists of columns that are associated with F.
- $\mathbf{X}_3$ consists of columns that are associated with effects that contain F.

The SWEEP operator applied on the augmented matrix $\mathbf{Z'WZ}$ is used to find the Type II sum of squares for each effect. The order of sweeping is determined by the "contained" relationship between the effect being tested and all other effects specified in the design.

Once the ordering is defined, the Type II sum of squares and its hypothesis matrix **L** can be obtained by a procedure similar to that used for the Type I sum of squares.

Constructing a Hypothesis Matrix

A hypothesis matrix **L** for the effect F has the form

$$\mathbf{L} = \left(\mathbf{0} \quad \mathbf{C X_2' W^{\frac{1}{2}} M_1 W^{\frac{1}{2}} X_2} \quad \mathbf{C X_2' W^{\frac{1}{2}} M_1 W^{\frac{1}{2}} X_3} \right)$$

where

$$\mathbf{M_1} = \mathbf{I} - \mathbf{W^{\frac{1}{2}} X_1 (X_1' W X_1)^* X_1 W^{\frac{1}{2}}}$$

$$\mathbf{C} = (\mathbf{X_2' W^{\frac{1}{2}} M_1 W^{\frac{1}{2}} X_2})^*$$

$\mathbf{A}^*$ is a g_2 generalized inverse of a symmetric matrix **A**.

Type III Sum of Squares and Hypothesis Matrix

The Type III sum of squares for an effect F can best be described as the sum of squares for F adjusted for effects that do not contain it, and *orthogonal* to effects (if any) that contain it.

Constructing a Hypothesis Matrix

A Type III hypothesis matrix **L** for an effect F is constructed using the following steps:

1. The design matrix **X** is reordered such that the columns can be grouped in three parts as described in the Type II approach. Compute $\mathbf{H} = (\mathbf{X'WX})^{*}\mathbf{X'WX}$. Notice that the columns of **H** can also be partitioned into three parts: the columns corresponding to effects not containing F, the columns corresponding to the effect F, and the columns corresponding to the effects containing F (if any).
2. The columns of those effects not containing F (except F) are *set to 0* by means of the row operation. That is:
   a) For each of those columns that is not already 0, fix any nonzero element in that column and call this nonzero element the pivot element.
   b) Divide the row that corresponds to the pivot element by the value of the pivot element itself.
   c) Use row operations to introduce zeros to all other nonzero elements (except the pivot element itself) in that column.
   d) Set the whole row containing the pivot element to 0. The column and the row corresponding to this pivot element should now be 0.
   e) Continue the process for the next column that is not 0 until all columns corresponding to those effects that do not contain F are 0.

3. For each column associated with effect F, find a nonzero element, use it as pivot, and perform the Gaussian elimination method as described in a, b, and c of step 2. After all such columns are processed, remove all of the 0 rows from the resulting matrix. If there is no column corresponding to effects containing F (which is the case when F contains all other effects), the matrix just constructed is the Type III hypothesis matrix for effect F. If there are columns corresponding to effects that contain F, continue with step 4.
4. The rows of the resulting matrix in step 3 can now be categorized into two groups. In one group, the columns corresponding to the effect F are all 0; call this group of rows G_0. In the other group, those columns are nonzero; call this group of rows G_1. Notice that the rows in G_0 form a generating basis for the effects that contain F. Transform the rows in G_1 such that they are orthogonal to any rows in G_0.

Calculating the Sum of Squares

Once a hypothesis matrix is constructed, the corresponding sum of squares can be calculated by $(\mathbf{L}\hat{\mathbf{B}})'(\mathbf{LGL}')^{*}\,\mathbf{L}\hat{\mathbf{B}}$.

Type IV Sum of Squares and Hypothesis Matrix

A hypothesis matrix **L** of a Type IV sum of squares for an effect F is constructed such that for each row of **L**, the columns corresponding to effect F are *distributed equitably* across the columns of effects containing F. Such a distribution is affected by the availability and the pattern of the nonmissing cells.

Constructing a Hypothesis Matrix

A Type IV hypothesis matrix **L** for effect F is constructed using the following steps:

1. Perform steps 1, 2, and 3 as described above for the Type III sum of squares.
2. If there are no columns corresponding to the effects containing F, the resulting matrix is a Type IV hypothesis matrix for effect F. If there are columns corresponding to the effects containing F, the following step is needed.
3. First, notice that each column corresponding to effect F represents a level in F. Moreover, the values in these columns can be viewed as the coefficients of a contrast for comparing different levels in F. For each row, the values of the columns corresponding to the effects that contain F are based on the values in that contrast. The final hypothesis matrix **L** consists of rows with nonzero columns corresponding to effect A. For each row with nonzero columns corresponding to effect F:
   a) If the value of any column (or level) corresponding to effect F is 0, set to 0 all values of columns corresponding to effects containing F *and* involving that level of F.
   b) For columns (or levels) of F that have nonzero values, count the number of times that those levels occur in one or more common levels of the other effects. This count will be based on the availability of the nonmissing cells in the data. Then set each column corresponding to an effect that contains F *and* involves that level of F to the value of the column that corresponds to that level of F divided by the count.
   c) If any value of the column corresponding to an effect that contains F *and* involves a level (column) of F is undetermined, while the value for that level (column) of F is nonzero, set the value to 0 and claim that the hypothesis matrix created may not be unique.

Calculating the Sum of Squares

Once a hypothesis matrix is constructed, the corresponding sum of squares can be calculated by $(\mathbf{L}\hat{\mathbf{B}})'(\mathbf{LGL}')^{*}\,\mathbf{L}\hat{\mathbf{B}}$. The corresponding degrees of freedom for this test is the row rank of the hypothesis matrix.