The General Linear Model: Theory

1.0 Introduction

In the discussion of multiple regression, we used the following equation to express the linear model for a single dependent variable:

$$\mathbf{y} = \mathbf{X}\theta + \mathbf{u} \tag{1.0}$$

where **y** is a column vector of the dependent variable, **X** is a matrix of the independent variables (with a column of 1's in the first column), θ is a vector of parameters and **u** is a vector of prediction errors. Strictly speaking, **X** is called a **design matrix** of independent variables, including the column of 1's for the intercept.

The mathematical equation for a univariate ANOVA is identical to that in Equation 1.0. The only difference is that what constitutes the "independent variables," or to put it another way, the elements of the design matrix is less obvious in ANOVA than it is in regression. Let us return to the example of the one-way ANOVA using therapy as the ANOVA factor. Instead of thinking of one independent variable, let us think of this as four different independent variables, one for each level of therapy. Call these variables X_1, X_2, X_3, and X_4. Those observations in the first level, the Control therapy, will score 1 on the first variable, X_1; all other observations score 0 on this variable. All those in the second therapy score 1 on X_2; 0, otherwise. And so on. These transformed variables now become the elements of the design matrix, **X**.

There are now five columns in **X**. The first is a column of 1's for the intercept. The second through fifth columns are the values of X_1 through X_4. The first parameter in θ is the intercept, the second parameter is the *b* coefficient for variable X_1, the third is the *b* coefficient for X_2, etc. We can now perform a multiple regression of the dependent variable, *Y*, on this design matrix. This regression is, in effect, identical to a one-way ANOVA (with one tiny exception that will be discussed later).

Hence, regression and ANOVA are mathematical equivalent. Furthermore, the analysis of covariance takes a similar form. Suppose that we would also like to control for age in the therapy example. We could take the same design matrix and add a column to the right of it and place the values of age in that column. We would then augment the vector θ by adding an additional parameter to the bottom of it--the regression coefficient for age. The fundamental equation, $\mathbf{y} = \mathbf{X}\theta + \mathbf{u}$, still holds. The only difference is that matrix **X** and vector θ got larger.

Now, let us complicate the example a bit more by adding a second dependent variable. We can do this by tacking on an additional column vector to the right-hand side of vector **y**, making it a matrix that we will denote as **Y**. We can then tack on another vector to the right side of θ, making it a matrix Θ. Finally, tack on another vector to **u**. The whole equation now looks like this:

$$\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \\ Y_{31} & Y_{32} \\ \vdots & \vdots \\ Y_{N1} & Y_{11} \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & X_{13} & X_{14} & age_1 \\ 1 & X_{21} & X_{22} & X_{23} & X_{24} & age_2 \\ 1 & X_{31} & X_{32} & X_{33} & X_{34} & age_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{N1} & X_{N1} & X_{N1} & X_{N1} & age_N \end{pmatrix} \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \\ \theta_{31} & \theta_{32} \\ \theta_{41} & \theta_{42} \\ \theta_{51} & \theta_{52} \\ \theta_{61} & \theta_{62} \end{pmatrix} + \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \\ U_{31} & U_{32} \\ \vdots & \vdots \\ U_{N1} & U_{N2} \end{pmatrix}$$

By continuing to add dependent variables, we would continue to add columns to matrices **Y**, Θ, and **U**. By adding more dependent variables, in either the form of ANOVA factors or continuous variables, we merely add columns to matrix X and rows to matrix Θ. Hence, matrices continue to get bigger as the problem becomes larger, but the fundamental equation stays the same. That equation is

$$\mathbf{Y} = \mathbf{X}\Theta + \mathbf{U} \tag{1.1}$$

and it is the equation for the **general linear model**. As we have developed this equation, ordinary multiple regression, multivariate multiple regression, the analysis of variance (ANOVA), the multivariate analysis of variance (MANOVA), the analysis of covariance (ANCOVA), and the multivariate analysis of covariance (MANCOVA) are all specific cases of the general linear model.

1.1 Hypothesis testing in the General Linear Model

Return to Equation 1.1 and postmultiply both sides of the equation by the transpose of matrix **Y**. This gives

$$\mathbf{YY}' = \mathbf{X}\Theta\mathbf{Y}' + \mathbf{UY}'$$

The null hypothesis states that all the elements of matrix Θ are 0. Consequently, all the elements of the matrix product $\mathbf{X}\Theta\mathbf{Y}'$ will be a large matrix filled with 0s. Consequently, we can write the equation

$$H_0 : \mathbf{X}\Theta\mathbf{Y}' = \mathbf{0}. \tag{1.2}$$

In English, this equation says, "the null hypothesis states (H_0:) that the product of the design matrix for the independent variables (**X**), the matrix of parameter estimates (Θ), and the transpose of the design matrix for the dependent variables (**Y**) is a matrix of 0s (**0**)."

The notation used in Equation 1.2 originated from the standard notation of multiple regression. At the risk of some confusion, notation will now be changed to agree with that in SAS documentation. Let **L** = **X**, β = Θ, and **M** = $\mathbf{Y}'$. Equation (1.2) now becomes

$$H_0 : L\beta\mathbf{M} = \mathbf{0} \tag{1.3}$$

Equation 1.3 is identical to Equation 1.2. Now, however,

L = design matrix for the independent variables.
β = matrix of parameters.
M = design matrix for the dependent variables.

The power of Equation 1.3 is that it *allows tests of hypotheses about linear combinations of the independent variables, linear combinations of the dependent variables, and linear combinations of both independent and dependent variables*. To restate this in more prosaic terms, the hypothesis testing in general linear models allows us to create new independent variables and/or new dependent variables from the original independent and dependent variables and then test hypotheses about the new variables.

1.2 Hypothesis Testing of Independent Variables: Contrast Coding

We begin examining hypothesis testing by focusing on the independent variables only. An example will be helpful in this case. .Suppose a research randomly assigned patients to four therapies to test therapeutic efficacy for some disorder. The first therapy is the standard therapy used in most clinics; the remaining three conditions are promising experimental therapies. Three different measures are used to assess outcome: (1) the Symptom Index, (SI), a measure of the number and intensity of symptoms present; (2) the Social Functioning measure (SF), a measure interpersonal activity; and (3) an Occupational Adjustment scale (OA), a measure of occupational functioning. All three measures are scored so that high scores denote better functioning. This is a one-way ANOVA design with three dependent variables and with therapy as the ANOVA factor.

To make matters simple, consider the typical hypothesis test for a univariate ANOVA with SI as the only the dependent measure. Algebraically, the model express the observed means for the four groups as a linear function of an overall mean and a deviation from the overall mean, or,

$$\bar{y}_1 = \mu + \alpha_1$$
$$\bar{y}_2 = \mu + \alpha_2$$
$$\bar{y}_3 = \mu + \alpha_3$$
$$\bar{y}_4 = \mu + \alpha_4$$

There are five unknown parameters, μ and the four αs. Thus, the vector of parameters is:

$$\beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix}$$

Because there is only a single dependent variable, matrix M becomes a (1 by 1) matrix containing the value of 1. Hence, it need not be written. Matrix **L** takes the form

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The reason that the first column in this matrix contains 0 is that we are uninterested in the overall mean, μ, and want to test whether the αs are 0. Hence, writing Equation 1.3 gives

$$H_0: \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

By performing the matrix multiplication, one can verify that the usual, one-way ANOVA in the case test the hypothesis that the four αs are equal to 0[1].

A significant *F* statistic tells us that it is unlikely the four groups were drawn from a single normal distribution on SI. It does not tell us which group(s) differ from which other group(s), nor does it tell us the direction of mean differences. Clearly, however, the intent of this research is to establish whether the experimental therapies perform better than the standard, control therapy and then to establish which of the experimental therapies work best. The simple *F* statistic from a one-way ANOVA does not answer these questions.

We can answer the pertinent research questions, however, by transforming the levels of the independent variables into "new" levels that directly address the hypothesis.

---

[1] I have taken liberties by allowing the parameters to be μ and the 4 αs because this approach is easy to understand. However, this approach has five parameters to predict four group means. Hence, there is no mathematical solution. One generally estimates an intercept or constant (μ) and three of the αs. It is for this reason that there are three degrees of freedom for the statistical test. This is also the reason why software programs like SAS will give messages saying that there is not a unique solution to the parameter estimates and that the printed solution is one of many alternatives.

Here, one expects that the mean SI score for the first group will be lower than the average mean for the last three groups. Thus, the *null hypothesis* states that the mean for the first group will equal the average mean for the last three groups. Expressing the null hypothesis algebraically gives:

$$H_0: \quad \bar{y}_1 = \frac{\bar{y}_2 + \bar{y}_3 + \bar{y}_4}{3}$$

For simplicity's sake we assume equal number of patients in the four groups. Now write the means in terms of μ and the αs instead of $\bar{Y}$:

$$H_0: \quad \mu + \alpha_1 = \frac{\mu + \alpha_2 + \mu + \alpha_3 + \mu + \alpha_4}{3}$$

Multiply each side by 3 and rearranging gives

$$0\mu + 3\alpha_1 - 1\alpha_2 - 1\alpha_3 - 1\alpha_4 = 0$$

The remaining trick is to express this in matrix notation:

$$\begin{pmatrix} 0 & 3 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = 0$$

This equation is identical in form to the hypothesis testing equation given in 1.3. Because there is a single dependent variable, matrix **M** is simply a (1 by 1) matrix consisting of the number 1. Hence, **M** does not need to be written. Matrix β in 1.3 is the column vector of μ and the αs. Finally, matrix **L** in 1.3 consists of the row vector (0, 3, -1, -1, -1).

Those of you familiar with contrast codes should immediately recognize that the vector **L** is nothing more than the vector of contrast codes! In fact, that's all that it is. We've gone through this is somewhat gory detail in order to show all the steps for developing contrast codes. They are:

(1) State the substantive hypothesis.
(2) State the null hypothesis for the substantive hypothesis.
(1) Write matrix (or vector) β.
(4) Write the algebraic model for the null hypothesis in terms of the parameters in matrix β

(5) Algebraically reduce the model in (4) until you get an expression with the parameters on the left and 0 on the right. The coefficients of the parameters become the elements in **L**. It is traditional, and indeed necessary for some software, to express these coefficients as integers.

Repeating these steps for the second hypothesis gives:

(1) Substantive hypothesis: there are some differences among the three experimental therapies.

(2) Null hypothesis: there are no differences among the three experimental therapies.

(3) β^t is the vector $\begin{pmatrix}\mu & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4\end{pmatrix}$.

(4) The null hypothesis states

$$\bar{Y}_2 = \bar{Y}_3 = \bar{Y}_4$$

so

$$\mu + \alpha_2 = \mu + \alpha_3 = \mu + \alpha_4.$$

(6) Begin reduction of the equation:

$$\mu + \alpha_2 = \mu + \alpha_3 = \mu + \alpha_4.$$
$$\alpha_2 = \alpha_3 = \alpha_4$$

Which can be written as two separate equations

$$\alpha_2 - \alpha_3 = 0$$
$$\alpha_2 - \alpha_4 = 0$$

Add these two equations together

$$\alpha_2 - \alpha_3 + \alpha_2 - \alpha_4 = 0$$

or

$$2\alpha_2 - \alpha_3 - \alpha_4 = 0.$$

Now add in parameters μ and α_1 by multiplying each by 0:

$$0\mu + 0\alpha_1 + 2\alpha_2 - \alpha_3 - \alpha_4 = 0.$$

This is the **L** vector or the vector of contrast codes.

The number of degrees of freedom for any contrast equals the number of rows in the **L** matrix. In the examples given above, two different hypotheses were tested, one after the other. Hence, each hypothesis test will have one degree of freedom associated with the numerator of the *F* ratio. Instead of sequential tests of individual hypotheses, it is possible to test two or more hypothesis simultaneously. For example, suppose, for

example, that the researcher performing this study was the one who actually developed the third experimental therapy. Suppose that the made the following hypotheses: (1) the fourth group will improve significantly more than the control group; (2) the fourth group will improve significantly greater than the average of the two other experimental therapies. An appropriate **L** matrix would be

$$\mathbf{L} = \begin{pmatrix} -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & -1 & 2 \end{pmatrix}$$

Here the researcher would have a single *F* test with two degrees of freedom to obtain an overall test of his/her idea. There are both advantages and disadvantages to this approach. The advantage is that if the researcher is indeed correct, then the test will be powerful. The disadvantage is that if only one of the hypotheses is true, then the overall test may fail to reveal differences and cannot determine which of the hypotheses is false.

There are two types of contrasts. In an **orthogonal contrast**, the rows of the L matrix are uncorrelated. Thus the matrix $\mathbf{LL}^t$ will be diagonal. The advantage of an orthogonal contrast is that each contrast is independent of each other contrast. When there are *k* levels to an ANOVA factor and there are *k* - 1 orthogonal contrasts (the maximum number of permissible contrasts), then the sums of squares for the ordinary ANOVA factor and the sum of sums of the squares for each contrast are equal. Thus if there are three degrees of freedom for the ANOVA factor, orthogonal contrasts break those 3 degrees of freedom into three independent hypotheses tests, each with one degree of freedom. Not only does this often increase power, but it also provides meaningful insight into where mean differences lie among the groups.

The second type of contrast is **nonorthogonal**. Here, matrix $\mathbf{LL}^t$ has at least one off-diagonal element that is not 0. Because the contrasts are correlated, there is some loss of inference analogous to the situation in multiple regression where there are correlations among the independent variables. The sums of squares for (*k* - 1) nonorthogonal contrasts will not add to the sum of squares for the ANOVA factor.

1.3 Hypothesis Testing of Dependent Variables: Transformations

Equation 1.3 is the general equation for multivariate hypothesis tests in MANOVA. It differs from the univariate test in two ways. First, Equation 1.3 contains the matrix **M**; in univariate tests, this matrix is a scalar equal to 1 and thus is not written. Second, in univariate tests, β will always be a column vector. In multivariate tests, β will have as many columns as there are variables. The first column will have the parameters for the first dependent variable, the second column will have the parameters for the second dependent variable, the third will have the parameters for the third dependent variable, and so on. In our example, we have three dependent variables, SI, SF, and OA. The β matrix would look like this

$$\beta = \begin{pmatrix} \mu_1 & \mu_2 & \mu_3 \\ \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ \alpha_{41} & \alpha_{42} & \alpha_{42} \end{pmatrix}$$

Here μ_1 is the overall mean for SI, the first variable, μ_2, the overall mean for the second variable, etc. For the αs, the first subscript denotes the group and the second denotes the variable. Thus, the MANOVA model for the mean of the second group on OA, the third variable, would be $\bar{y}_{23} = \mu_3 + \alpha_{23}$, and the mean of the fourth group on the second variable (SF) would be $\bar{y}_{42} = \mu_2 + \alpha_{42}$.

The purpose of the M matrix is to create one or more new dependent variables from the original dependent variables and then perform the analysis on the new dependent variables. This will be called a **transformation** of the dependent variables. The general form of the transformation is

$$\mathbf{Y}^* = \mathbf{M}^t \mathbf{Y}$$

where $\mathbf{Y}^*$ are the new dependent variables, $\mathbf{Y}$ are the old dependent variables, and $\mathbf{M}^t$ is the transpose of the $\mathbf{M}$ matrix.

For example, suppose that we were interested in the overall level of functioning among the patients in this study. We could create a new variable in this by summing the variables SI, SF, and OA. The equation would be

$$Y^* = (1 \quad 1 \quad 1) \begin{pmatrix} SI \\ SF \\ OA \end{pmatrix}$$

If the simple one-way ANOVA were performed on these data, then the equation

$$H_0 : \mathbf{L}\beta\mathbf{M} = \mathbf{0}$$

reduces to

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 & \mu_2 & \mu_3 \\ \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 + \mu_2 + \mu_3 \\ \alpha_{11} + \alpha_{12} + \alpha_{13} \\ \alpha_{21} + \alpha_{22} + \alpha_{23} \\ \alpha_{31} + \alpha_{32} + \alpha_{33} \\ \alpha_{41} + \alpha_{42} + \alpha_{43} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} \alpha_{11} + \alpha_{12} + \alpha_{13} \\ \alpha_{21} + \alpha_{22} + \alpha_{23} \\ \alpha_{31} + \alpha_{32} + \alpha_{33} \\ \alpha_{41} + \alpha_{42} + \alpha_{43} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

In short, this tests whether the deviations from the means summed over the three variables are equal to 0.

1.4 Standard Types of Contrasts and Transformations

The purpose for contrasts of independent variables and transformations of dependent variables is to make sense of the data (in exploratory analysis) and to rigorously test hypotheses (in hypothesis testing analysis). Consequently, there are no strict rules about contrasts or transformations that apply to every type of GLM problem. Nevertheless, there are some standard types of contrasts or transformations that tend to be used more often than others. The standard types are:

DIFFERENCE: A difference contrast compares a standard level against all other levels of an ANOVA factor. In the therapy example, each of the three experimental therapies may be contrasted with the control therapy. Ignoring the coefficient for μ, The contrast code matrix would be

$$\mathbf{L} = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

For dependent variables, a difference transformation would compare a standard dependent variable against each of the other dependent variables. This would be useful when the standard is a baseline measure and the other dependent variables are measures of change over time.

HELMERT: A Helmert contrast compares a level to the mean of subsequent levels. Like the polynomial, this is most useful when the ANOVA levels or the MANOVA dependent variables have some order. For the drug example given above, a Helmert contrast for the four groups would be

$$\mathbf{L} = \begin{pmatrix} 3 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

MEAN: A mean contrast or transformation compares the mean for one level (or a dependent variable) against the mean of the other levels (or dependent variables).

POLYNOMIAL: Polynomial contrasts and transformations assume that the levels of the ANOVA factors (or the ordering of the dependent variables) have some natural order. The first contrast is a linear contrast, the second is a quadratic contrast, the third, a cubic contrast, and so on. An example would be a one-way ANOVA where a control group received a placebo drug and three experimental groups received increasing doses of an active drug. The contrast matrix testing for the linear and quadratic effects would be

$$\mathbf{L} = \begin{pmatrix} -2 & -1 & 1 & 2 \\ -1 & 1 & 1 & -1 \end{pmatrix}$$

Polynomial transformations are very useful in repeated measures designs.

PROFILE: A profile contrast or transformation compares each level to the adjacent level (or each dependent variable to the adjacent dependent variable). An example of a profile transform would be the following M matrix for the three dependent variables in the therapy example:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}$$

Profile transformations are particularly useful when one wishes to determine whether group differences on a set of dependent variables are due merely to elevation (e.g., group 1 just has overall higher scores than group 2) or to shape (e.g., group1 responds significantly higher on dependent variable 1 than it does on dependent variable 2).

Be warned that there is no standard nomenclature for contrasts and transformations. Some software programs, for example, call a difference transformation a “contrast” transformation, and other packages may call it a difference transformation but compare dependent variables to the last dependent variable. It is important to always consult the software documentation before performing transformations and contrasts.