The Matrix-Based Approach to the General Linear Model

We have seen that both regression and analysis of variance models are examples of the general linear model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i,$$

and we can now notice that an individual observation from this model can be expressed as an inner product of the variable vector (usually including a '1' to multiply the intercept) and the parameter vector, plus the error:

$$Y_i = \mathbf{x}_i'\beta + \varepsilon_i,$$

where $\mathbf{x}_i'$ is of dimension 1 x $(k+1)$ and β is of dimension $(k+1)\,\mathrm{x}\,1$. The model for the entire sample can then be expressed as a matrix equation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}.$$

The matrix $\mathbf{X}$ is often called the model matrix, or for analysis of variance models, the design matrix. The usual assumptions that the errors ε have mean 0, constant variance σ^2, and are independent can be stated in matrix form:

$$E(\boldsymbol{\varepsilon}) = \mathbf{0},\ V(\boldsymbol{\varepsilon}) \equiv E[(\boldsymbol{\varepsilon}-E(\boldsymbol{\varepsilon}))(\boldsymbol{\varepsilon}-E(\boldsymbol{\varepsilon}))'] = E[(\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon})'] = \sigma_\varepsilon^2\mathbf{I}_n,$$

and if we further assume that the errors ε are multivariately normally distributed, then since

$$\boldsymbol{\mu} \equiv E(\mathbf{y}) = E(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta}+E(\boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta},$$

we can write that $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma_\varepsilon^2\mathbf{I}_n)$.

Model matrices for regression, dummy regression, one-way ANOVA

Discussion of distinct patterns that occur in model matrices for ANOVA applications, and distinctions between model matrices $\mathbf{X}$, full-rank model matrices $\mathbf{X}_F$, and row-basis model matrices $\mathbf{X}_B$. Since a row-basis matrix is a square matrix of full rank, we can invert it to solve for the relationships between model parameters and cell means via $\boldsymbol{\beta}_F = \mathbf{X}_B^{-1}\boldsymbol{\mu}$.

Linear Contrasts

Although researchers often focus on the global ANOVA null hypothesis that there is no effect of a factor, it is possible to specify comparisons among means (linear contrasts) that are of specific interest, most conveniently using the matrix $\mathbf{X}_B^{-1}$. The comparisons, called linear contrasts, can be constructed either to examine sub-hypotheses of the main null hypothesis, or if constructed to be orthogonal when the design is balanced, they can yield independent sub-tests of the main null hypothesis.

The Least-Squares Estimator

Given the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, we can define the least-squares estimator $\mathbf{b}$ as the solution to the residual sum of squares:

$$\begin{aligned} S(\mathbf{b}) &= \sum E_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \\ &= \mathbf{y}'\mathbf{y} - (2\mathbf{y}'\mathbf{X})\mathbf{b} + \mathbf{b}'(\mathbf{X}'\mathbf{X})\mathbf{b}. \end{aligned}$$

To minimize $S(\mathbf{b})$ we calculate its vector partial derivative

$$\frac{\partial S(\mathbf{b})}{\partial \mathbf{b}} = \mathbf{0} - 2\mathbf{X}'\mathbf{y}+2\mathbf{X}'\mathbf{X}\mathbf{b},$$

which is set to zero to yield the normal equations

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}.$$

If the matrix $\mathbf{X}$ is of full column rank $(k + 1)$ then $\mathbf{X}'\mathbf{X}$ is nonsingular, so we can invert it to obtain the least-squares estimator:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

For $\mathbf{X}$ to be of full column rank $(k + 1)$, we must have at least as many observations n as coefficients $k+1$, and the $k+1$ columns of $\mathbf{X}$ must be linearly independent. The text points out that since the second partial derivative of $S(\mathbf{b})$ is positive-definite (when $\mathbf{X}'\mathbf{X}$ is nonsingular), we know that the solution $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ minimizes $S(\mathbf{b})$.