Appendix 3

Further Topics in Matrix Algebra and Linear Models

This appendix builds on Chapter 8, presenting additional results from matrix algebra and linear models. We start by introducing two useful matrix transforms, generalized inverses (for solving singular systems of equations) and the square root of a matrix (for obtaining a set of uncorrelated variables). These results are then used for a formal derivation of several properities of generalized least-squares (GLS) estimators. We next examine how linear model sums of squares can be written as quadratic forms and how these sums of squares are used in formal hypothesis testing. We conclude with two additional topics, equivalent linear models (which allow calculations for one model to be performed on a potentially much simpler model) and a brief introduction to matrix derivatives.

GENERALIZED INVERSES AND SOLUTIONS TO SINGULAR SYSTEMS OF EQUATIONS

Linear systems of equations are ubiquitous in quantitative genetics and we have presented solutions for such systems by assuming that the appropriate matrices are nonsingular, and hence can be inverted. However, in the real world of large, complex, and unbalanced designs, the existence of an inverse is by no means guaranteed. Consider the solution of the matrix equation $\mathbf{y} = \mathbf{Ax}$ for the unknown vector $\mathbf{x}$. If $\mathbf{A}$ is a square and nonsingular, then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ is the unique solution. However, what happens if $\mathbf{A}$ is singular or is nonsquare? In this case either the system has no solution and is said to be **inconsistent** or else there are an infinite number of solutions. An example of an inconsistent system is

$$x_1 + x_2 = 1$$
$$x_1 + x_2 = 2$$

which cannot be satisfied by any $(x_1,\, x_2)$. Likewise, a system with an infinite number of solutions is

$$x_1 + x_2 = 1$$
$$x_1 + x_2 = 1$$

which has a line of solutions of the form $x_2 = 1 - x_1$ for arbitrary x_1. While these two simple systems can be solved by inspection, a more systematic approach is required for arbitary systems. This is provided by using **generalized inverses**.

Generalized Inverses

Suppose a matrix $\mathbf{A}^-$ exists such that

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A} \tag{A3.1}$$

where $\mathbf{A}$ is $p \times q$ and $\mathbf{A}^-$ is $q \times p$. Premultiplying both sides of the equation $\mathbf{Ax} = \mathbf{y}$ by $\mathbf{AA}^-$ gives

$$\mathbf{AA}^-\mathbf{Ax} = \mathbf{Ax} = \mathbf{AA}^-\mathbf{y}$$

and hence

$$\mathbf{A}(\mathbf{x} - \mathbf{A}^-\mathbf{y}) = \mathbf{0}$$

implying that, if the system is consistent, a solution is

$$\mathbf{x} = \mathbf{A}^-\mathbf{y} \tag{A3.2}$$

Given the analogy with the inverse of a nonsingular square matrix, a matrix $\mathbf{A}^-$ satisfying Equation A3.1 is called a generalized inverse (also **g-inverse**, **conditional inverse**) of $\mathbf{A}$. Unless $\mathbf{A}$ is nonsingular, Equation A3.1 does not define a unique matrix, so we refer to $\mathbf{A}^-$ as *a* generalized inverse instead of *the* generalized inverse. A unique generalized inverse, the **Moore-Penrose inverse**, can be obtained by imposing three additional conditions: $\mathbf{A}^-\mathbf{AA}^- = \mathbf{A}^-$, $(\mathbf{AA}^-)^T = \mathbf{AA}^-$, and $(\mathbf{A}^-\mathbf{A})^T = \mathbf{A}^-\mathbf{A}$. However, for our purposes any $\mathbf{A}^-$ satisfying Equation A3.1 is sufficient. Methods for computing generalized inverses are found in Henderson (1984a). More detailed treatment of the properties of generalized inverses are given by Dhrymes (1978), Searle (1982), Pringle and Rayner (1971), and Rao and Mitra (1971), and we summarize some of these results below.

Consistency and Solutions to Consistent Systems

When dealing with linear models for complex designs, it is not immediately clear if the resulting OLS/GLS equations have solutions. Generalized inverses provide a check of consistency, and hence of whether a system of equations has any solutions. A linear system $\mathbf{Ax} = \mathbf{y}$ is consistent if and only if

$$\mathbf{AA}^-\mathbf{y} = \mathbf{y} \tag{A3.3}$$

Given a consistent system, all solutions have the form

$$\mathbf{x} = \mathbf{A}^-\mathbf{y} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{c} \tag{A3.4}$$

where $\mathbf{c}$ is an arbitrary $q \times 1$ column vector. For example, taking $\mathbf{c} = \mathbf{0}$ recovers Equation A3.2, while if $\mathbf{A}^{-1}$ exists, then $\mathbf{I} - \mathbf{A}^{-1}\mathbf{A} = \mathbf{0}$ and the solution $x = \mathbf{A}^{-1}\mathbf{y}$

is unique. To see that any expression of the form of Equation A3.4 is a solution, note that

$$\begin{aligned}\mathbf{Ax} &= \mathbf{A}(\mathbf{A}^-\mathbf{y} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{c}) \\ &= \mathbf{AA}^-\mathbf{y} + (\mathbf{A} - \mathbf{AA}^-\mathbf{A})\mathbf{c} = \mathbf{y} + (\mathbf{A} - \mathbf{A})\mathbf{c} \\ &= \mathbf{y}\end{aligned}$$

which follows from Equations A3.3 and A3.1, respectively.

Example 1. Consider the following system of equations

$$\begin{aligned}x_1 + 2x_2 + 3x_3 &= 5 \\ 2x_1 + x_2 + 2x_3 &= 6\end{aligned}$$

which can be written in matrix form as $\mathbf{Ax} = \mathbf{y}$, with

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

The matrix

$$\mathbf{A}^- = \begin{pmatrix} -11/26 & 9/13 \\ 4/13 & -3/13 \\ 7/26 & -1/13 \end{pmatrix}$$

satisfies $\mathbf{AA}^-\mathbf{A} = \mathbf{A}$ and thus is a generalized inverse of $\mathbf{A}$. Matrix multiplication shows that $\mathbf{AA}^- = \mathbf{I}$, implying $\mathbf{AA}^-\mathbf{y} = \mathbf{y}$. Thus, Equation A3.3 is satisfied and this system of equations is consistent for any $\mathbf{y}$. One solution is $\mathbf{x} = \mathbf{A}^-\mathbf{y}$, or

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -11/26 & 9/13 \\ 4/13 & -3/13 \\ 7/26 & -1/13 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \frac{1}{26} \begin{pmatrix} 53 \\ 4 \\ 23 \end{pmatrix}$$

More generally, since

$$\mathbf{I} - \mathbf{A}^-\mathbf{A} = \begin{pmatrix} 1/26 & 2/13 & -3/26 \\ 2/13 & 8/13 & -6/13 \\ -3/26 & -6/13 & 9/26 \end{pmatrix}$$

then from Equation A3.4, any solution to this system of equations has the form

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{1}{26} \begin{pmatrix} 53 \\ 4 \\ 23 \end{pmatrix} + \begin{pmatrix} 1/26 & 2/13 & -3/26 \\ 2/13 & 8/13 & -6/13 \\ -3/26 & -6/13 & 9/26 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

which reduces to

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{1}{26}\begin{pmatrix} 53 \\ 4 \\ 23 \end{pmatrix} + c \cdot \begin{pmatrix} 1 \\ 4 \\ -3 \end{pmatrix}$$

where c is an arbitrary constant. Substitution shows this to be a solution.

Although an infinite number of solutions exists when $\mathbf{A}$ is singular, particular linear combinations (or **contrasts**) of the elements of $\mathbf{x}$ may have unique values. For example, consider the system $x_1 + x_2 = 1$. Here there are an infinite number of solutions for (x_1, x_2), but only a single solution, 1, for the contrast $x_1 + x_2$.

Consider some linear combination $\mathbf{b}^T\mathbf{x}$. If the vector of constants $\mathbf{b}$ satisfies

$$\mathbf{b}^T\mathbf{A}^-\mathbf{A} = \mathbf{b}^T \tag{A3.5a}$$

then $\mathbf{b}^T\mathbf{x}$ has a unique solution given by

$$\mathbf{b}^T\mathbf{x} = \mathbf{b}^T\mathbf{A}^-\mathbf{y} \tag{A3.5b}$$

To see this, note that Equation A3.4 gives the general solution as

$$\begin{aligned} \mathbf{b}^T\mathbf{x} &= \mathbf{b}^T(\mathbf{A}^-\mathbf{y} + [\mathbf{I} - \mathbf{A}^-\mathbf{A}]\mathbf{c}) \\ &= \mathbf{b}^T\mathbf{A}^-\mathbf{y} + (\mathbf{b}^T\mathbf{I} - \mathbf{b}^T\mathbf{A}^-\mathbf{A})\mathbf{c} \\ &= \mathbf{b}^T\mathbf{A}^-\mathbf{y} + (\mathbf{b}^T - \mathbf{b}^T)\mathbf{c} \\ &= \mathbf{b}^T\mathbf{A}^-\mathbf{y} \end{aligned}$$

which is independent of the arbitrary vector $\mathbf{c}$. Likewise, a vector of contrasts $\mathbf{Bx}$ has a unique solution $\mathbf{BA}^-\mathbf{y}$, provided $\mathbf{B}$ satisfies $\mathbf{BA}^-\mathbf{A} = \mathbf{B}$

Example 2. Consider the system of equations from Example 1. Is there a unique solution for the two linear contrasts $c_1 = x_2 - 4x_1$ and $c_2 = x_3 + 3x_1$? In matrix form,

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} x_2 - 4x_1 \\ x_3 + 3x_1 \end{pmatrix} = \mathbf{Bx}$$

where

$$\mathbf{B} = \begin{pmatrix} -4 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Using the generalized inverse for $\mathbf{A}$ from Example 1, matrix multiplication shows that

$$\mathbf{BA^-A} = \begin{pmatrix} -4 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} = \mathbf{B}$$

Hence, the matrix version of Equation A3.5b gives the unique solution for this vector of contrasts as

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \mathbf{BA^-y} = \begin{pmatrix} -4 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} -11/26 & 9/13 \\ 4/13 & -3/13 \\ 7/26 & -1/13 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} -8 \\ 7 \end{pmatrix}$$

To see that this solution is indeed unique, note that we can rearrange the contrast equations to obtain $x_2 = c_1 + 4x_1$ and $x_3 = c_2 - 3x_1$. Substituting into the original set of equations (Example 1),

$$x_1 + 2x_2 + 3x_3 = x_1 + 2(c_1 + 4x_1) + 3(c_2 - 3x_1) = 2c_1 + 3c_2 = 5$$
$$2x_1 + x_2 + 2x_3 = 2x_1 + (c_1 + 4x_1) + 2(c_2 - 3x_1) = c_1 + 2c_2 = 6$$

so that the original set of three equations and three unknowns reduces to a two equation-two unknown system. In matrix form this is

$$\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

Since the coefficient matrix is invertible, there is a unique solution for this pair of contrasts ($c_1 = -8$ and $c_2 = 7$).

Estimability of Fixed Factors

The above results have implications for the estimation of (fixed) factors in the general linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. Recall that the OLS solution for a vector $\boldsymbol{\beta}$ of fixed effects is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ (Chapters 8, 26). If the design matrix $\mathbf{X}$ has **full column rank** (all columns of $\mathbf{X}$ are independent), $(\mathbf{X}^T\mathbf{X})^{-1}$ exists and the OLS solution for $\boldsymbol{\beta}$ is unique. However, when $(\mathbf{X}^T\mathbf{X})$ is singular (and hence does not have a unique inverse), it is not possible to obtain unique OLS estimates for all the fixed factors in a model. For example, suppose β_1 indicates a sex effect and β_2 indicates the effect of a particular diet. If the design is such that all females use this diet, we do not have separate information on both sex and diet effects and hence can only estimate $\beta_1 + \beta_2$ rather than being able to estimate both β_1 and β_2 separately.

A linear combination of factors $\mathbf{b}^T\boldsymbol{\beta}$ is said to be **estimable** for a given design matrix $\mathbf{X}$ if there exists some column vector $\mathbf{a}$ that satisfies

$$E(\mathbf{a}^T\mathbf{y}) = \mathbf{b}^T\boldsymbol{\beta} \qquad \text{(A3.6a)}$$

Estimability thus implies that there is some linear combination $\mathbf{a}^T\mathbf{y}$ of the original data whose expected value equals the desired linear combination of factors. Since $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, this definition implies that $\mathbf{b}^T\boldsymbol{\beta}$ is estimable if there exists a column vector $\mathbf{a}$ that satisfies $E(\mathbf{a}^T\mathbf{y}) = \mathbf{a}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{b}^T\boldsymbol{\beta}$, implying $(\mathbf{a}^T\mathbf{X} - \mathbf{b}^T)\boldsymbol{\beta} = \mathbf{0}$, or

$$\mathbf{X}^T\mathbf{a} = \mathbf{b} \tag{A3.6b}$$

An alternative (and equivalent) condition is that $\mathbf{b}$ satisfies

$$\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^-(\mathbf{X}^T\mathbf{X}) = \mathbf{b}^T \tag{A3.6c}$$

Henderson (1984a) gives other equivalent conditions. Equation A3.6c implies that if $\mathbf{X}^T\mathbf{X}$ is nonsingular, all linear combinations of $\boldsymbol{\beta}$ are estimable. Note that Equation A3.6c is identical to the condition given by Equation A3.5a (taking $\mathbf{A} = \mathbf{X}^T\mathbf{X}$), implying that these solutions are also unique estimates. If estimable, the OLS solution of the vector $\mathbf{b}^T\boldsymbol{\beta}$ given by

$$\text{OLS}(\mathbf{b}^T\boldsymbol{\beta}) = \mathbf{b}^T\left(\mathbf{X}^T\mathbf{X}\right)^-\mathbf{X}^T\mathbf{y} \tag{A3.6d}$$

is unique and independent of which generalized inverse is actually used.

Example 3. Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$, where

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{giving} \quad \mathbf{X}^T\mathbf{X} = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Note that $\mathbf{X}^T\mathbf{X}$ is singular, so we cannot obtain unique estimates of all three parameters. For this design matrix, are β_3, $\beta_1+\beta_2$, and β_1 estimable? These three combinations correspond to vectors of $\mathbf{b}^T = (0,0,1)$, $(1,1,0)$, and $(1,0,0)$, respectively. For the first two $\mathbf{b}$ vectors, we can find a vector $\mathbf{a}$ that satisfies $\mathbf{X}^T\mathbf{a} = \mathbf{b}$, viz.,

$$\mathbf{X}^T\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{X}^T\begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

so that, from Equation A3.6b, these two linear combinations, β_3 and $(\beta_1+\beta_2)$, are estimable. However, since

$$\mathbf{X}^T\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1+a_2 \\ a_1+a_2 \\ a_3 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

β_1 is not estimable as $a_1 + a_2$ cannot simultaneously equal zero and one, and hence there exists no vector $\mathbf{a}$ that satisfies $\mathbf{X}^T\mathbf{a} = \mathbf{b}$ for this particular $\mathbf{X}$ and $\mathbf{b}$.

THE SQUARE ROOT OF A MATRIX

The concept of the square root of a symmetric nonsingular matrix provides another useful matrix tool for the analysis of linear models. In particular, using the square root of the covariance matrix transforms a vector of correlated variables into a new vector of variables with covariance matrix $\mathbf{I}$, implying that the transformed variables are uncorrelated with unit variance.

Consider a symmetric nonsingular matrix $\mathbf{V}$ and define $\mathbf{V}^{1/2}$ as the matrix satisfying

$$\mathbf{V}^{1/2}\mathbf{V}^{1/2} = \mathbf{V} \tag{A3.7a}$$

In effect, $\mathbf{V}^{1/2}$ is the square root of a matrix, in that, when squared, we recover $\mathbf{V}$. Denoting the inverse of $\mathbf{V}^{1/2}$ as $\mathbf{V}^{-1/2}$, we have the following properties

$$\mathbf{V}^{-1/2}\mathbf{V}^{1/2} = \mathbf{I}, \quad \mathbf{V}^{-1/2}\mathbf{V}^{-1/2} = \mathbf{V}^{-1}, \quad \text{and} \quad \mathbf{V}^{-1/2}\mathbf{V} = \mathbf{V}^{1/2} \tag{A3.7b}$$

Likewise, both $\mathbf{V}^{1/2}$ and its inverse are symmetric.

Suppose the random vector $\mathbf{y}$ has covariance matrix $\mathbf{V}$ and consider the new vector $\mathbf{z} = \mathbf{V}^{-1/2}\mathbf{y}$. Recalling Equation 8.21b, the resulting covariance matrix for $\mathbf{z}$ becomes

$$\text{Var}(\mathbf{z}) = \mathbf{V}^{-1/2}\text{Var}(\mathbf{y})\mathbf{V}^{-1/2} = \mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \mathbf{I} \tag{A3.8}$$

Thus, the transformed variables have unit variance and are uncorrelated. Suppose $\mathbf{y}$ is an $n \times 1$ column vector with $\mathbf{y} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{V})$. It follows that

$$\mathbf{z} = \mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) \sim \text{MVN}(\mathbf{0}, \mathbf{I})$$

so that $z_i \sim \text{N}(0, 1)$, and hence the transformed variables are independent unit normals. Thus,

$$\begin{aligned}(\mathbf{y} - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) &= (\mathbf{y} - \boldsymbol{\mu})^T\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) \\ &= \mathbf{z}^T\mathbf{z} \\ &= \sum_{i=1}^{n} z_i^2 \sim \chi_n^2 \end{aligned} \tag{A3.9}$$

The last step follows by recalling that the sum of n squared unit normal random variables follows a χ^2 distribution with n degrees of freedom (Appendix 5). Thus when $\mathbf{y}$ is multivariate normal, the quadratic form $(\mathbf{y}-\boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ follows a χ^2 distribution. As we will see shortly, Equation A3.9 is the basis for goodness-of-fit tests of linear models.

DERIVATION OF THE GLS ESTIMATORS

One important application of the square root of a matrix is that it allows us to obtain generalized least-squares (GLS) estimators from ordinary least-squares (OLS) estimators. Suppose the linear model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \qquad \text{with } \mathbf{e} \sim (0, \mathbf{R}\,\sigma_e^2)$$

Premultiplying both sides by $\mathbf{R}^{-1/2}$ gives

$$\mathbf{z} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{f} \qquad \text{with } \mathbf{f} \sim (0, \mathbf{I}\,\sigma_e^2)$$

where

$$\mathbf{z} = \mathbf{R}^{-1/2}\mathbf{y}, \qquad \mathbf{Z} = \mathbf{R}^{-1/2}\mathbf{X}, \qquad \mathbf{f} = \mathbf{R}^{-1/2}\mathbf{e}$$

OLS can be applied to this model since the transformed residuals are uncorrelated and homoscedastic. Thus, GLS estimates are obtained from the OLS solution by substituting

$$\mathbf{z} = \mathbf{R}^{-1/2}\mathbf{y} \ \text{ for } \ \mathbf{y}, \qquad \mathbf{Z} = \mathbf{R}^{-1/2}\mathbf{X} \ \text{ for } \ \mathbf{X}, \qquad \mathbf{f} = \mathbf{R}^{-1/2}\mathbf{e} \ \text{ for } \ \mathbf{e} \tag{A3.10}$$

Substituting into the OLS solutions (Equation 8.33a) gives the GLS estimate of $\boldsymbol{\beta}$ as

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \Big((\,\mathbf{X}^T\mathbf{R}^{-1/2}\,)\,(\,\mathbf{R}^{-1/2}\mathbf{X}\,)\Big)^{-1} \,(\,\mathbf{X}^T\mathbf{R}^{-1/2}\,)\,(\,\mathbf{R}^{-1/2}\mathbf{y}\,) \\ &= \Big(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\Big)^{-1}\,\mathbf{X}^T\mathbf{R}^{-1}\mathbf{y}\end{aligned}$$

Likewise, substituting into the OLS covariance expression (Equation 8.33b) gives the resulting covariance matrix for the GLS estimates as

$$\text{Var}(\,\widehat{\boldsymbol{\beta}}\,) = \Big(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\Big)^{-1}\,\sigma_e^2$$

If the residuals follow a multivariate normal distribution, $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{V})$, and $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ is indeed the correct model, then $\mathbf{y} - \widehat{\mathbf{y}} \sim \text{MVN}(\mathbf{0}, \mathbf{V})$ and it follows from Equation A3.9 that

$$(\mathbf{y} - \widehat{\mathbf{y}})^T\,\mathbf{V}^{-1}\,(\mathbf{y} - \widehat{\mathbf{y}}) \sim \chi^2 \tag{A3.11a}$$

The degrees of freedom for the χ^2 distribution equal the number of observations minus the number of estimated parameters. Equation A3.11a provides a χ^2 test for the goodness-of-fit of a particular linear model. If $\mathbf{V}$ is a diagonal matrix, then

$$(\mathbf{y} - \widehat{\mathbf{y}})^T\,\mathbf{V}^{-1}\,(\mathbf{y} - \widehat{\mathbf{y}}) = \sum_{i=1}^{n} \frac{(\,y_i - \widehat{y}_i\,)^2}{V_{ii}} \sim \chi^2 \tag{A3.11b}$$

Similar modifications extend a number of other OLS results into GLS results (Table A3.1).

Table A3.1 Summary of useful results for the general linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, under ordinary least-squares (OLS) and generalized least-squares (GLS) assumptions about the distribution of residuals.

	OLS	GLS
Assumed distribution of residuals	$\mathbf{e} \sim (\mathbf{0}, \sigma_e^2\,\mathbf{I})$	$\mathbf{e} \sim (\mathbf{0}, \mathbf{V})$
Least-squares estimator of $\boldsymbol{\beta}$	$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$
Var($\widehat{\boldsymbol{\beta}}$)	$(\mathbf{X}^T\mathbf{X})^{-1}\sigma_e^2$	$(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$
Predicted values, $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$	$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	$\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$
Var($\widehat{\mathbf{y}}$)	$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma_e^2$	$\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T$
Chi-square goodness-of-fit statistic (assuming $\mathbf{e} \sim$ MVN)	$\chi^2 = \sum_{i=1}^{n} \frac{(\,y_i - \widehat{y}_i\,)^2}{\sigma_e^2}$	$\chi^2 = (\mathbf{y} - \widehat{\mathbf{y}})^T\mathbf{V}^{-1}(\mathbf{y} - \widehat{\mathbf{y}})$

QUADRATIC FORMS AND SUMS OF SQUARES

The analysis of linear models relies very heavily on sums of squares, which can be expressed in matrix notation as quadratic forms. To introduce the reader to the machinery used to work with sums of squares, we first present expressions for the mean and variance of a quadratic form, and then express linear model sums of squares as quadratic forms.

Moments of Quadratic Forms

When $\mathbf{x}$ is a vector of random variables, the quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is a scalar random variable. If $\mathbf{x}$ has mean $\boldsymbol{\mu}$ and (nonsingular) covariance matrix $\mathbf{V}$, Equation 8.22 gives the expected value of this quadratic form as

$$E(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \text{tr}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \tag{A3.12a}$$

where the **trace** of a square matrix, $\text{tr}(\mathbf{M}) = \sum M_{ii}$, is the sum of its diagonal elements. Further, if $\mathbf{x} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{V})$, then as shown in Searle (1971), the variance of the quadratic form has a fairly simple form,

$$\sigma^2(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\,\text{tr}\,(\mathbf{A}\mathbf{V}\mathbf{A}\mathbf{V}) + 4\boldsymbol{\mu}^T\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu} \tag{A3.12b}$$

The Sample Variance Expressed as a Quadratic Form

As an introduction to expressing sums of squares as quadratic forms, consider the sample variance for n observations,

$$\text{Var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Define the **unit matrix** $\mathbf{J}_{n\times k}$ as an $n \times k$ matrix in which every element is unity, e.g.,

$$\mathbf{J}_{n\times 1} = \left.\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right\}n\,, \qquad \mathbf{J}_{2\times 3} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Likewise, define the matrix

$$\mathbf{N} = \frac{1}{n-1}\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right) = \frac{1}{n-1}\begin{pmatrix} 1-1/n & -1/n & \cdots & -1/n \\ -1/n & 1-1/n & \cdots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \cdots & 1-1/n \end{pmatrix} \tag{A3.13a}$$

where $\mathbf{J}$ is $n \times n$. Noting that

$$\mathbf{N}\mathbf{x} = \frac{1}{n-1}\left(\mathbf{x} - \frac{1}{n}\mathbf{J}\mathbf{x}\right) = \frac{1}{n-1}\begin{pmatrix} x_1 - \overline{x} \\ \vdots \\ x_n - \overline{x} \end{pmatrix} \tag{A3.13b}$$

it follows that

$$\mathbf{x}^T\mathbf{N}\mathbf{x} = \text{Var}(x) \tag{A3.14a}$$

To see this, observe that

$$\begin{aligned} \mathbf{x}^T\mathbf{N}\mathbf{x} &= \frac{1}{n-1}\,(\,x_1 \quad \cdots \quad x_n\,)\begin{pmatrix} x_1 - \overline{x} \\ \vdots \\ x_n - \overline{x} \end{pmatrix} \\ &= \frac{1}{n-1}\sum_{i=1}^{n} x_i(\,x_i - \overline{x}\,) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \overline{x}\sum_{i=1}^{n} x_i\right) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}\,)^2 = \text{Var}(x) \end{aligned} \tag{A3.14b}$$

Example 4. Since we have expressed Var(x) as a quadratic form, we can use Equation A3.12a to compute its expected value and Equation A3.12b (under the assumption of normality) to compute its sampling variance. If $\mathbf{x} \sim (\boldsymbol{\mu}, \mathbf{V})$, the expected value of Var(x) is

$$E[\,\text{Var}(x)\,] = E(\mathbf{x}^T\mathbf{N}\mathbf{x}) = \text{tr}(\mathbf{N}\mathbf{V}) + \boldsymbol{\mu}^T\mathbf{N}\boldsymbol{\mu}$$

To compute this expression, first note from Equation A3.14b that

$$\boldsymbol{\mu}^T\mathbf{N}\boldsymbol{\mu} = \frac{1}{n-1}\sum_{i=1}^{n}(\,\mu_i - \overline{\mu}\,)^2$$

Likewise, from Equation A3.13b

$$\mathbf{N}\mathbf{V} = \frac{\mathbf{V}}{n-1} - \frac{\mathbf{J}\mathbf{V}}{n(n-1)}$$

which has diagonal elements

$$(\mathbf{N}\mathbf{V})_{ii} = \frac{1}{n-1}\left(\sigma^2(z_i) - \frac{\sum_j \sigma(z_i, z_j)}{n}\right)$$

After some simplification, we have

$$\text{tr}(\mathbf{N}\mathbf{V}) = \sum_{i=1}^{n}(\mathbf{N}\mathbf{V})_{ii} = \frac{1}{n}\sum_{i=1}^{n}\sigma^2(z_i) - \frac{2}{n(n-1)}\sum_{i<j}\sigma(z_i, z_j)$$

Putting these results together gives

$$E[\text{Var}(x)] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2(z_i) - \frac{2}{n(n-1)}\sum_{i<j}\sigma(z_i, z_j) + \frac{1}{n-1}\sum_{i=1}^{n}(\,\mu_i - \overline{\mu}\,)^2$$

where $\overline{\mu} = \sum \mu_i/n$. In the simple situation where all observations have the same mean and variance ($\mu_i = \mu$, $\sigma^2(z_i) = \sigma^2$) and are uncorrelated, this reduces to

$$E[\,\text{Var}(x)\,] = \sigma^2$$

Turning now to the sample variance of Var(x), if we are willing to assume that $\mathbf{x}$ is multivariate normal, then from Equation A3.12b,

$$\sigma^2[\text{Var}(x)] = \sigma^2(\mathbf{x}^T\mathbf{N}\mathbf{x}) = 2\,\text{tr}\,[\mathbf{N}\mathbf{V}\mathbf{N}\mathbf{V}] + 4\boldsymbol{\mu}^T\mathbf{N}\mathbf{V}\mathbf{N}\boldsymbol{\mu}$$

If, for example, $\mathbf{V} = \sigma^2\,\mathbf{I}$ (the x_i are uncorrelated with common variance), then

$$\mathbf{NVNV} = \sigma^4\,\mathbf{NN} = \frac{\sigma^4}{(n-1)^2}\left(\mathbf{I} - \frac{1}{n}\,\mathbf{J}_{n\times n}\right)\left(\mathbf{I} - \frac{1}{n}\,\mathbf{J}_{n\times n}\right)$$
$$= \frac{\sigma^4}{(n-1)^2}\left(\mathbf{I} - \frac{2}{n}\,\mathbf{J}_{n\times n} + n^{-2}\,\mathbf{J}_{n\times n}\,\mathbf{J}_{n\times n}\right)$$

The *ij*th element in $\mathbf{J}_{n\times n}\,\mathbf{J}_{n\times n}$ is n, giving $\mathbf{J}^2_{n\times n} = n\,\mathbf{J}_{n\times n}$. Hence, the *i*th diagonal element of $\mathbf{NVNV}$ is

$$\frac{\sigma^4}{(n-1)^2}\left(1 - \frac{2}{n} + n^{-2}n\right) = \frac{\sigma^4}{n(n-1)}$$

giving tr$(\mathbf{NVNV}) = \sigma^4/(n-1)$. When all of the means are equal, it follows that $\mathbf{N}\boldsymbol{\mu} = \mathbf{0}$ and the second term in Equation A3.12b vanishes, giving

$$\sigma^2[\,\mathrm{Var}(x)\,] = \frac{2\sigma^4}{n-1}$$

Sums of Squares Expressed as Quadratic Forms

In the same fashion that we decomposed total variance into genetic and phenotypic components (Chapters 3–7), we can decompose the total variance of a response vector $\mathbf{y}$ into the variance accounted for by the linear model and the remaining (error or residual) variance. This is typically done by considering the sums of squares, with the **total sum of squares** ($\mathrm{SS_T}$) being the sum of two components, the **error** (or **residual**) **sum of squares** ($\mathrm{SS_E}$) and the **model sum of squares** ($\mathrm{SS_M}$),

$$\mathrm{SS_T} = \mathrm{SS_M} + \mathrm{SS_E}$$

The total sum of squares measures the total variability in the data, while the model sum of squares measures the amount of variation accounted for by the linear model. As noted in our discussions of univariate regression in Chapter 3, the fraction of total variance explained by a linear model is given by the **coefficient of determination**,

$$r^2 = \frac{\mathrm{SS_M}}{\mathrm{SS_T}} = 1 - \frac{\mathrm{SS_E}}{\mathrm{SS_T}} \tag{A3.15}$$

The sums of squares have different forms under OLS and GLS. Under OLS, the residuals are assumed to be independent with common variance σ^2_e. In this case, each observation / residual is weighted equally, and the total sum of squares is simply

$$\mathrm{SS_T} = \sum_{i=1}^{n}(\,y_i - \overline{y}\,)^2$$

Sums of squares can be expressed as a quadratic form of the vector of observations $\mathbf{y}$, allowing the use of Equations 3A.12a,b to obtain their expectations and variances. Recalling Equation A3.14b and A3.13a,

$$\mathrm{SS_T} = \mathbf{y}^T \left(\mathbf{I} - \frac{1}{n}\,\mathbf{J} \right) \mathbf{y} \tag{A3.16a}$$

where $\mathbf{J}$ is $n \times n$.

Now consider the error sum of squares

$$\mathrm{SS_E} = \sum_{i=1}^{n} (\, y_i - \widehat{y}_i\,)^2 = \sum_{i=1}^{n} \widehat{e}_i^{\,2}$$

Since $\widehat{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}}$ and $\widehat{\mathbf{y}} = \mathbf{Xb} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$, we have

$$\mathrm{SS_E} = \widehat{\mathbf{e}}^T\widehat{\mathbf{e}}, \quad \text{where} \quad \widehat{\mathbf{e}} = \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T \right] \mathbf{y} \tag{A3.16b}$$

Expanding this expression and noting that $\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \mathbf{I}$, this simplifies to

$$\mathrm{SS_E} = \mathbf{y}^T \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T \right] \mathbf{y} \tag{A3.16c}$$

Finally, the model sum of squares is the difference between the total and error sums of squares,

$$\mathrm{SS_M} = \mathrm{SS_T} - \mathrm{SS_E} = \mathbf{y}^T \left[\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T - \frac{1}{n}\,\mathbf{J} \right] \mathbf{y} \tag{A3.16d}$$

Note that

$$\mathrm{SS_M} = \sum_{i=1}^{n} (\, \widehat{y}_i - \overline{y}\,)^2$$

so that (for OLS) the model sum of squares is the sum of squared deviations of the predicted values from the overall mean.

The sums of squares under generalized least-squares (GLS) are slightly different, as we have to correct for heteroscedasticity and/or the lack of independence among the residuals. Assume that the residuals have covariance matrix $\sigma_e^2\,\mathbf{R}$. From Equation A3.10, $\mathbf{y}$ is replaced by $\mathbf{R}^{-1/2}\mathbf{y}$ and $\mathbf{X}$ is replaced by $\mathbf{R}^{-1/2}\mathbf{X}$ in the above OLS expressions for sums of squares. Hence, the total sum of squares for GLS becomes

$$\begin{aligned}\mathrm{SS_T} &= \mathbf{y}^T\mathbf{R}^{-1/2}\left(\mathbf{I} - \frac{1}{n}\,\mathbf{J} \right)\mathbf{R}^{-1/2}\mathbf{y} \\ &= \mathbf{y}^T \left[\mathbf{R}^{-1} - \frac{1}{n}\,\mathbf{R}^{-1/2}\mathbf{J}\mathbf{R}^{-1/2} \right] \mathbf{y}\end{aligned} \tag{A3.17a}$$

Likewise, the error sum of squares becomes

$$\begin{aligned} \text{SS}_{\text{E}} &= \widehat{\mathbf{e}}^T \mathbf{R}^{-1} \widehat{\mathbf{e}} \\ &= \mathbf{y}^T \left[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X} \left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{R}^{-1} \right] \mathbf{y} \end{aligned} \tag{A3.17b}$$

and the model sum of squares becomes

$$\text{SS}_{\text{M}} = \mathbf{y}^T \left[\mathbf{R}^{-1}\mathbf{X} \left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{R}^{-1} - \frac{1}{n}\, \mathbf{R}^{-1/2} \mathbf{J} \mathbf{R}^{-1/2} \right] \mathbf{y} \tag{A3.17c}$$

TESTING HYPOTHESES ABOUT LINEAR MODELS

Since sums of squares are very closely related to the variances accounted for by the various components of a particular linear model, it should not be surprising that hypothesis testing is based on the sums of squares. Such hypothesis tests can be quite involved, especially if we are evaluating the various components of a complex model. Here we consider the simplest case of testing the fit of the total model to the data.

If the residuals are multivariate-normally distributed with

$$\mathbf{e} \sim \text{MVN}(\mathbf{0}, \sigma_e^2\, \mathbf{I}) \quad \text{for OLS;} \qquad \mathbf{e} \sim \text{MVN}(\mathbf{0}, \sigma_e^2\, \mathbf{R}) \quad \text{for GLS}$$

then (recalling Equation A3.11a and A3.17b), $\text{SS}_{\text{E}}/\sigma_e^2$ is the sum of squared unit normals and hence is χ^2-distributed. In particular, with n observations and p estimated parameters,

$$\frac{\text{SS}_{\text{E}}}{\sigma_e^2} \sim \chi^2_{n-p} \tag{A3.18}$$

as a degree of freedom is lost for each estimated model parameter.

Suppose we have n observations and wish to compare two linear models, a **full model** fitting p parameters and a **reduced model** which uses only a subset ($q < p$) of the parameters in the full model. Do the additional $p - q$ fitted parameters provide a significant increase in the amount of variation accounted for by the model? Let SS_{E_f} and SS_{E_r} denote the appropriate (OLS or GLS) error sums of squares for the full and reduced models. Under the null hypothesis (that the full model provides the same fit as the reduced model), the difference in error sums of squares ($\text{SS}_{\text{E}_r} - \text{SS}_{\text{E}_f}$) is distributed as constant (σ_e^2) times a χ^2_{p-q}. Likewise, from Equation A3.18, $\text{SS}_{\text{E}_f} \sim \sigma_e^2\, \chi^2_{n-p}$. Recalling the definition of the F distribution (Appendix 5), it follows that

$$\frac{\left(\text{SS}_{\text{E}_r} - \text{SS}_{\text{E}_f}\right) / \,(p-q)}{\text{SS}_{\text{E}_f} / \,(n-p)} = \left(\frac{n-p}{p-q} \right) \left(\frac{\text{SS}_{\text{E}_r}}{\text{SS}_{\text{E}_f}} - 1 \right) \tag{A3.19}$$

is distributed as $F_{p-q,n-p}$ under the null hypothesis of no improved fit.

For example, we can ask if a particular linear model accounts for a significant fraction of the variation in y by considering that model versus the simplest reduced model $y_i = \mu + e_i$. It is easily seen that the least-squares solution for μ is $\overline{y}$ for OLS and the weighted mean for GLS, giving $\mathrm{SS}_{E_r} = \mathrm{SS}_\mathrm{T}$. Since the number of parameters in the reduced model is $q = 1$, the test for whether a particular linear model accounts for a significant amount of the variation is

$$\left(\frac{n-p}{p-1}\right)\left(\frac{\mathrm{SS}_\mathrm{T}}{\mathrm{SS}_{\mathrm{E}_f}} - 1\right) = \left(\frac{n-p}{p-1}\right)\left(\frac{r^2}{1-r^2}\right) \tag{A3.20}$$

where r^2 is the coefficient of determination for the full model (Equation A3.15). This test statistic follows an $F_{p-1,n-p}$ distribution.

EQUIVALENT LINEAR MODELS

Two linear models are said to be **equivalent** if they have the same mean vector $E(\mathbf{y})$ and covariance matrix $\sigma(\mathbf{y},\mathbf{y})$. The utility of equivalent models is that the parameters of one model can always be expressed as linear combinations of the parameters of any equivalent model. Hence, by choosing an appropriate equivalent model, one can often greatly simplify computations. An example of this approach is the reduced animal model of Quaas and Pollak (1980) discussed in Chapter 26. Likewise, Equation 26.23, for estimating the BLUP values of dominance effects as a function of estimated breeding values, also follows from using equivalent models. Additional examples from BLUP are given by Henderson (1985c). Our purpose here is to briefly introduce the use and construction of equivalent models.

Consider two different mixed linear models, both using the same vector $\mathbf{y}$ of observations but with different assumed vectors of fixed ($\boldsymbol{\beta}$ vs. $\boldsymbol{\beta}_*$) and random ($\mathbf{u}$ and $\mathbf{e}$ vs. $\mathbf{u}_*$ and $\mathbf{e}_*$) effects. Model 1 is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad \text{where} \quad \mathbf{u} \sim (\mathbf{0},\mathbf{G}) \quad \text{and} \quad \mathbf{e} \sim (\mathbf{0},\mathbf{R})$$

while model 2 is

$$\mathbf{y} = \mathbf{X}_*\boldsymbol{\beta}_* + \mathbf{Z}_*\mathbf{u}_* + \mathbf{e}_*, \quad \text{where} \quad \mathbf{u}_* \sim (\mathbf{0},\mathbf{G}_*) \quad \text{and} \quad \mathbf{e}_* \sim (\mathbf{0},\mathbf{R}_*)$$

Recalling our treatment of general mixed linear models (Chapter 26), Equation 26.2 implies that for model 1,

$$\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta},\mathbf{V}), \quad \text{where} \quad \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$$

while for model 2,

$$\mathbf{y} \sim (\mathbf{X}_*\boldsymbol{\beta}_*,\mathbf{V}_*), \quad \text{where} \quad \mathbf{V}_* = \mathbf{Z}_*\mathbf{G}_*\mathbf{Z}_*^T + \mathbf{R}_*$$

Thus, these two models are equivalent if

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_*\boldsymbol{\beta}_* \tag{A3.21a}$$

and $\mathbf{V} = \mathbf{V}_*$, or

$$\mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} = \mathbf{Z}_*\mathbf{G}_*\mathbf{Z}_*^T + \mathbf{R}_* \tag{A3.21b}$$

Equations A3.21a,b provide the framework for constructing equivalent models, and hence obtaining models that are potentially easier to analyze. Consider the situation where our interest is in the prediction of random effects and we wish to obtain an equivalent model that considers the same fixed effects but uses a different vector of random effects. (For example, instead of considering a vector of both parental and offspring breeding values, we might simply consider the vector of parental breeding values, using the parental estimates to subsequently estimate the breeding values in their offspring.) If the original model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \qquad \text{where} \quad \mathbf{u} \sim (0, \mathbf{G}), \quad \text{and} \quad \mathbf{e} \sim (0, \mathbf{R})$$

an equivalent model using *any* vector of random effects $\mathbf{u}_* \sim (0, \mathbf{G}_*)$ is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_*\mathbf{u}_* + \mathbf{e}_*, \qquad \text{where} \quad \mathbf{u}_* \sim (0, \mathbf{G}_*), \quad \text{and} \quad \mathbf{e}_* \sim (0, \mathbf{R}_*)$$

Since for these models to be equivalent, we require that $\mathbf{V} = \mathbf{V}_*$, it immediately follows from Equation A3.21b that the covariance matrix for the vector of new residual values, $\mathbf{e}_*$, is given by

$$\mathbf{R}_* = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^T - \mathbf{Z}_*\mathbf{G}_*\mathbf{Z}_*^T \tag{A3.22}$$

Given an estimate of $\mathbf{u}_*$, an estimate of $\mathbf{u}$ can be directly obtained, as parameters of a linear model can always be expressed as linear combinations of the parameters of any equivalent model. In this case, given the BLUP estimate ($\widehat{\mathbf{u}}_*$) of $\mathbf{u}_*$, the BLUP estimate of $\mathbf{u}$ is given by

$$\widehat{\mathbf{u}} = \mathbf{C}\mathbf{G}^{-1}\widehat{\mathbf{u}}_* \tag{A3.23}$$

where $\mathbf{C}$ is the covariance matrix between $\mathbf{u}_*$ and $\mathbf{u}$, and $\mathbf{G}$ is the covariance matrix associated with $\mathbf{u}$ (Henderson 1977b). This is just the linear regression of $\mathbf{u}_*$ on $\mathbf{u}$ (see Equation 8.27). Note that the vectors $\mathbf{u}_*$ and $\mathbf{u}$ can have different dimensionality, so that if $\mathbf{u}_*$ is $r \times 1$ and $\mathbf{u}$ is $q \times 1$, then $\mathbf{C}$ is an $r \times q$ matrix with $C_{ij} = \sigma(u_{*i}, u_j)$.

DERIVATIVES OF VECTORS AND MATRICES

Our final special topic in matrix algebra concerns the derivatives of vector- and matrix-valued functions, which we use rather extensively in Chapter 27. We present a few simple results here, and the reader is referred to Morrison (1976), Graham (1981), and Searle (1982) for more details. Consider first the simplest function of vector $\mathbf{x}$, namely the product of $\mathbf{x}$ and either a vector ($\mathbf{a}$) or matrix ($\mathbf{A}$) of constants. The derivatives of these functions with respect to the vector $\mathbf{x}$ become

$$\frac{\partial\,\mathbf{a}^T\mathbf{x}}{\partial\,\mathbf{x}} = \frac{\partial\,\mathbf{x}^T\mathbf{a}}{\partial\,\mathbf{x}} = \mathbf{a} \tag{A3.24a}$$

$$\frac{\partial\,\mathbf{A}\mathbf{x}}{\partial\,\mathbf{x}} = \mathbf{A}^T \tag{A3.24b}$$

Turning to quadratic forms, if $\mathbf{A}$ is symmetric, then

$$\frac{\partial\,\mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial\,\mathbf{x}} = 2\mathbf{A}\mathbf{x} \tag{A3.25a}$$

Three useful identities involving quadratic forms follow

$$\frac{\partial\,(\mathbf{a}-\mathbf{x})^T\mathbf{A}(\mathbf{a}-\mathbf{x})}{\partial\,\mathbf{x}} = -2\mathbf{A}(\mathbf{a}-\mathbf{x}) \tag{A3.25b}$$

$$\frac{\partial\,(\mathbf{a}-\mathbf{B}\mathbf{x})^T(\mathbf{a}-\mathbf{B}\mathbf{x})}{\partial\,\mathbf{x}} = -2\mathbf{B}^T(\mathbf{a}-\mathbf{B}\mathbf{x}) \tag{A3.25c}$$

$$\frac{\partial\,(\mathbf{a}-\mathbf{B}\mathbf{x})^T\mathbf{A}(\mathbf{a}-\mathbf{B}\mathbf{x})}{\partial\,\mathbf{x}} = -2\mathbf{B}^T\mathbf{A}(\mathbf{a}-\mathbf{B}\mathbf{x}) \tag{A3.25d}$$

Example 5. The OLS solution for a linear model is the value of $\boldsymbol{\beta}$ that minimizes the residual sum of squares given $\mathbf{y}$ and $\mathbf{X}$. In matrix form,

$$\sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e} = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})$$

Taking the derivative with respect to $\boldsymbol{\beta}$ and using Equation A3.25c (with $\mathbf{a}=\mathbf{y}$, $\mathbf{B}=\mathbf{X}$, and $\mathbf{x}=\boldsymbol{\beta}$) gives

$$\frac{\partial\,\mathbf{e}^T\mathbf{e}}{\partial\,\boldsymbol{\beta}} = \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\partial\,\boldsymbol{\beta}} = -2\mathbf{X}^T\,(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

Setting this equal to zero gives $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$, or

$$\boldsymbol{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

If $\mathbf{X}^T\mathbf{X}$ is singular, a generalized inverse is used instead.