

## OVERVIEW OF LINEAR MODELS

Linear models form the backbone of most estimation procedures in quantitative genetics and will be extensively used throughout the rest of this book. They are generally structured such that a vector of observations of one variable ( $y$ ) is modeled as a linear combination of other variables observed along with  $y$ . The remainder of this chapter introduces some of the basic tools and key concepts underlying the use of linear models. Advanced topics are examined in detail in Chapters 26 and 27, and further comments are given in Appendix 3.

In multiple regression, the commonest type of linear model, the predictor variables  $x_1, \dots, x_n$  represent observed values for  $n$  traits of interest. More generally, some or all of the predictor variables could be **indicator variables**, with values of 0 or 1 indicating whether an observation belongs in a particular category or grouping of interest. As an example, consider the half-sib design wherein each of  $p$  unrelated sires is mated at random to a number of unrelated dams and a single offspring is measured from each cross. The simplest model for this design is

$$y_{ij} = \mu + s_i + e_{ij}$$

where  $y_{ij}$  is the phenotype of the  $j$ th offspring from sire  $i$ ,  $\mu$  is the population mean,  $s_i$  is the **sire effect**, and  $e_{ij}$  is the residual error (the “noise” remaining in the data after the sire effect is removed). Although this is clearly a linear model, it differs significantly from the regression model described above in that while there are parameters to estimate (the sire effects  $s_i$ ), the only measured values are the  $y_{ij}$ . Nevertheless, we can express this model in a form that is essentially identical to the standard regression model by using  $p$  indicator (i.e., zero or one) variables to classify the sires of the offspring. The resulting linear model becomes

$$y_{ij} = \mu + \sum_{k=1}^p s_k x_{ik} + e_{ij}$$

where

$$x_{ik} = \begin{cases} 1 & \text{if sire } k = i \\ 0 & \text{otherwise} \end{cases}$$

By the judicious use of indicator variables, an extremely wide class of problems can be handled by linear models. Models containing only indicator variables are usually termed ANOVA (**analysis of variance**) models, while regression usually refers to models in which predictor variables can take on a continuous range of values. Both procedures are special cases of the **general linear model** (GLM), wherein each observation ( $y$ ) is assumed to be a linear function of  $p$  observed and/or indicator variables plus a residual error ( $e$ ),

$$y_i = \sum_{k=1}^p \beta_k x_{ik} + e_i \quad (8.32a)$$

where  $x_{i1}, \dots, x_{ip}$  are the values of the  $p$  predictor variables for the  $i$ th individual. For a vector of  $n$  observations, the GLM can be written in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (8.32b)$$

where the **design** or **incidence matrix**  $\mathbf{X}$  is  $n \times p$ , and  $\mathbf{e}$  is the vector of residual errors. It is important to note that  $\mathbf{y}$  and  $\mathbf{X}$  contain the observed values, while  $\boldsymbol{\beta}$  is a vector of parameters (usually called **factors** or **effects**) to be estimated.

**Example 8.** Suppose that three different sires used in the above half-sib design have two, one, and three offspring, respectively. This can be expressed in GLM form,  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$  with

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{31} \\ y_{32} \\ y_{33} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ s_1 \\ s_2 \\ s_3 \end{pmatrix}, \quad \text{and} \quad \mathbf{e} = \begin{pmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{31} \\ e_{32} \\ e_{33} \end{pmatrix}$$

Likewise, the multiple regression

$$y_i = \alpha + \sum_{j=1}^p \beta_j x_{ij} + e_i$$

can be written in GLM form with

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \text{and} \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

### Polynomial Regressions and Interaction Effects

The general linear model is extremely flexible, covering many models that at first appear nonlinear. The key here is that linearity refers to the **parameters** to be estimated, not the data being observed. For example, consider the **quadratic regression** of  $y$  on a single predictor variable  $x$ . The  $i$ th observation is assumed to be of the form

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

While quadratic terms appear, these are in the observed variable, *not* the parameters to be estimated  $(\alpha, \beta_1, \beta_2)$ . Expressed in GLM form,

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

This sort of reasoning extends to polynomial regressions of any order and to just about any transformation of the observed variables, e.g., we could consider  $\ln x$  or  $e^{-x}$ . Likewise the GLM can be used to account for **interaction effects**. For example, consider the model

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} \cdot x_{i2} + e_i$$

which in matrix form is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

with

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11} \cdot x_{12} \\ 1 & x_{21} & x_{22} & x_{21} \cdot x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1} \cdot x_{n2} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

The  $\beta_j$  for this model have a slightly different interpretation from partial regression coefficients. Here, with  $x_1$  held constant, a unit change in  $x_2$  changes the mean value of  $y$  by  $\beta_2 + \beta_3 \cdot x_1$ . Likewise with  $x_2$  held constant, a unit change in  $x_1$  changes the mean value by  $\beta_1 + \beta_3 \cdot x_2$ .

### Fixed vs. Random Effects

Linear models are based upon sets of variables that classify individuals into various groupings, often referred to as **factors** or **effects**. For example, suppose we have information on the sex of an individual, which diet it was raised on, and its age. These are the three factors for this analysis, and we can ask how much of the variation in a response variable is attributable to each factor individually and to interactions between the various factors (i.e., a sex-diet interaction not predicted by sex or diet alone).

There are two fundamentally different types of factors — **fixed** and **random**. The distinction between fixed and random effects is usually straightforward, but at times it can be extremely subtle. Consider the simple model in which a single factor takes on  $k$  discrete values. Then,  $y_{ij} = \mu + \beta_j + e_{ij}$ , where  $1 \leq j \leq k$ , so that  $y_{ij}$  is the  $i$ th observation at the  $j$ th value of the factor. Whether the factor is treated as fixed or random depends on how the  $k$  values of the factor are chosen. Under a random effects model, the  $k$  values are drawn at random from a probability

distribution with mean zero and unknown variance. In this case, our interest is usually in estimating the variance of this distribution. Conversely, we may decide on a fixed set of factor values in advance (such as males versus females, or  $k$  distinct diets). These are **fixed effects**, as the factor values are assumed to be fixed in advance of the analysis, i.e., there is no variance associated with their choice. Because the distinction between fixed and random effects lies in how we treat the underlying sampling distribution of factor values, situations arise in which one investigator assumes fixed factor values, while another regards them as random. For example, when  $k$  diets are assayed, diet is a random effect if we regard the treatments as a random sample from some universe of possible diets. Conversely, if we are interested in these  $k$  *particular* diets, then diet is a fixed effect.

General linear models are used for three rather different classes of estimation problems: estimating fixed effects, estimating the variance of random effects, and predicting random effects. When dealing with random effects, the literature usually refers to **predicting**, rather than estimating their values, to distinguish this from the estimation of fixed effects. The rest of this chapter is largely concerned with the estimation of fixed effects. Estimation of variance components dominates much of our discussion in Chapters 17-24 (which use ANOVA methods), leading up to the more advanced and flexible method of REML (restricted maximum likelihood) estimation of variance components discussed in Chapter 27. The ascertainment of values of random variables is of great concern in animal breeding, where the prediction of breeding values is a key issue. The most powerful approach here is BLUP (best linear unbiased prediction), which allows for **mixed** models with both fixed and random effects (Chapter 26).

**Example 9.** Consider the sire model,  $z_{ij} = \mu + s_i + e_{ij}$ , that we used to motivate the concept of indicator variables. If five sires are measured, we could regard these individuals as a random sample from a larger population of interest. With this interpretation, the sire effects are random effects drawn from a distribution with mean 0 (a non-zero mean is absorbed by the population mean  $\mu$  included in the model) and variance  $\sigma_s^2$ . Estimation of the sire variance is of interest as a means of estimating the additive genetic variance, since (ignoring complicating factors)  $\sigma_s^2 = \sigma_A^2/4$  (Chapter 7). In addition to estimating the additive genetic variance of the population from which the five sires were drawn, we may also wish to predict the specific sire values  $s_1, \dots, s_5$ , as for example, when the individuals with the highest sire values are being sought for future breeding. Conversely, if these five males are all the sires in our breeding program and we do not plan to introduce any outside sires, we could regard this group of five as the entire population of sires. In this case, sires are fixed effects, as there is no population that we wish to draw inferences on other than these five sires.

**ORDINARY LEAST SQUARES, OLS**

Estimates of the vector  $\beta$  for the general linear model are usually obtained by the method of least-squares, which uses the observations  $\mathbf{y}$  and  $\mathbf{X}$  and makes special assumptions about the covariance structure of the vector of residual errors  $\mathbf{e}$ . The method of **ordinary least squares** assumes that the residual errors are homoscedastic and uncorrelated, i.e.,  $\sigma^2(e_i) = \sigma_e^2$  for all  $i$ , and  $\sigma(e_i, e_j) = 0$  for  $i \neq j$ .

Let  $\mathbf{b}$  be an estimate of  $\beta$ , and denote the vector of  $y$  values predicted from this estimate by  $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$ , so that the resulting vector of residual errors is

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

The ordinary least-squares (OLS) estimate of  $\beta$  is the  $\mathbf{b}$  vector that minimizes the residual sum of squares,

$$\sum_{i=1}^n \hat{e}_i^2 = \hat{\mathbf{e}}^T \hat{\mathbf{e}} = (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b})$$

Taking derivatives, it can be shown that our desired estimate satisfies

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \tag{8.33a}$$

Under the assumption that the residual errors are uncorrelated and homoscedastic (i.e., the covariance matrix of the residuals is  $\sigma_e^2 \cdot \mathbf{I}$ ), the covariance matrix of the elements of  $\mathbf{b}$  is

$$\mathbf{V}_{\mathbf{b}} = (\mathbf{X}^T \mathbf{X})^{-1} \sigma_e^2 \tag{8.33b}$$

Hence, the OLS estimator of  $\beta_i$  is the  $i$ th element of the column vector  $\mathbf{b}$ , while the variance of this estimator is the  $i$ th diagonal element of the matrix  $\mathbf{V}_{\mathbf{b}}$ . Likewise, the covariance of this estimator with the OLS estimator for  $\beta_j$  is the  $ij$ th element of  $\mathbf{V}_{\mathbf{b}}$ .

If the residuals follow a multivariate normal distribution with  $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \sigma_e^2 \cdot \mathbf{I})$ , the OLS estimate is also the maximum-likelihood estimate. If  $\mathbf{X}^T \mathbf{X}$  is singular, Equations 8.33a,b still hold when a generalized inverse is used, although only certain linear combinations of fixed factors can be estimated (see Appendix 3 for details).

**Example 9.** Consider a univariate regression where the predictor and response variable both have expected mean zero, so that the regression passes through the origin. The appropriate model becomes

$$y_i = \beta x_i + e_i$$

With observations on  $n$  individuals, this relationship can be written in GLM form with  $\beta = \beta$  and design matrix  $\mathbf{X} = (x_1, x_2, \dots, x_n)^T$ , implying

$$\mathbf{X}^T \mathbf{X} = \sum_{i=1}^n x_i^2 \quad \text{and} \quad \mathbf{X}^T \mathbf{y} = \sum_{i=1}^n x_i y_i$$

Applying Equations 8.33a,b gives the OLS estimate of  $\beta$  and its sample variance (assuming the covariance matrix of  $\mathbf{e}$  is  $\mathbf{I} \cdot \sigma_e^2$ ) as

$$b = \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y} = \frac{\sum x_i y_i}{\sum x_i^2}, \quad \sigma^2(b) = \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \sigma_e^2 = \frac{\sigma_e^2}{\sum x_i^2}$$

This estimate of  $\beta$  differs from the standard univariate regression slope (Equation 3.14b) where the intercept value is not assumed to be equal to zero.

**Example 10.** Recall from Equation 8.10b that the vector of partial regression coefficients for a multivariate regression is defined to be  $\mathbf{b} = \mathbf{V}^{-1} \mathbf{c}$  (where  $\mathbf{V}$  is the estimated covariance matrix, and  $\mathbf{c}$  is the vector of estimated covariances between  $\mathbf{y}$  and  $\mathbf{z}$ ). Here we show that this expression is equivalent to the OLS estimator  $\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ . Using the notation from Example 8, for the  $i$ th individual we observe  $y_i$  and the values of  $p$  predictor variables,  $z_{i1}, \dots, z_{ip}$ . Since the regression satisfies  $\bar{y} = \alpha + \beta_1 \bar{z}_1 + \dots + \beta_p \bar{z}_p$ , subtracting the mean from each observation removes the intercept, with

$$y_i^* = (y_i - \bar{y}) = \beta_1(z_{i1} - \bar{z}_1) + \dots + \beta_p(z_{ip} - \bar{z}_p) + e_i$$

For  $n$  observations, the resulting linear model  $\mathbf{y}^* = \mathbf{X}\beta + \mathbf{e}$  has

$$\mathbf{y}^* = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} (z_{11} - \bar{z}_1) & \cdots & (z_{1p} - \bar{z}_p) \\ \vdots & \ddots & \vdots \\ (z_{n1} - \bar{z}_1) & \cdots & (z_{np} - \bar{z}_p) \end{pmatrix}$$

where  $z_{ij}$  is the value of character  $j$  in the  $i$ th individual. Partitioning the design matrix  $\mathbf{X}$  into  $p$  column vectors corresponding to the  $n$  observations on each of the  $p$  predictor variables gives

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p) \quad \text{where} \quad \mathbf{x}_j = \begin{pmatrix} z_{1j} - \bar{z}_j \\ z_{2j} - \bar{z}_j \\ \vdots \\ z_{nj} - \bar{z}_j \end{pmatrix}$$

giving the  $j$ th element of the vector  $\mathbf{X}^T \mathbf{y}^*$  as

$$\left( \mathbf{X}^T \mathbf{y}^* \right)_j = \mathbf{x}_j^T \mathbf{y}^* = \sum_{i=1}^n (y_i - \bar{y})(z_{ij} - \bar{z}_j) = (n-1) \text{Cov}(y, z_j)$$

and implying  $\mathbf{X}^T \mathbf{y}^* = (n - 1) \mathbf{c}$ . Likewise, the  $jk$ th element of  $\mathbf{X}^T \mathbf{X}$  is

$$\mathbf{x}_j^T \mathbf{x}_k = \sum_{i=1}^n (z_{ij} - \bar{z}_j)(z_{ik} - \bar{z}_k) = (n - 1) \text{Cov}(z_j, z_k)$$

implying  $\mathbf{X}^T \mathbf{X} = (n - 1) \mathbf{V}$ . Putting these results together gives

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}^* = \mathbf{V}^{-1} \mathbf{c}$$

showing that Equation 8.10b does indeed give the OLS estimates of the partial regression coefficients.

### 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  $\mathbf{y} = \mathbf{A}\mathbf{x}$  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

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 + x_2 &= 2 \end{aligned}$$

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

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 + x_2 &= 1 \end{aligned}$$

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 arbitrary 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

$$\mathbf{x} = \mathbf{A}^- \mathbf{y} \quad (\text{A3.2})$$

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

$$\mathbf{AA}^- \mathbf{y} = \mathbf{y} \quad (\text{A3.3})$$

Given a consistent system, all solutions have the form

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

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

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

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

$$\mathbf{b}^T \mathbf{A}^{-} \mathbf{A} = \mathbf{b}^T \quad (\text{A3.5a})$$

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

$$\mathbf{b}^T \mathbf{x} = \mathbf{b}^T \mathbf{A}^{-} \mathbf{y} \quad (\text{A3.5b})$$

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  $\mathbf{B}\mathbf{x}$  has a unique solution  $\mathbf{B}\mathbf{A}^{-} \mathbf{y}$ , provided  $\mathbf{B}$  satisfies  $\mathbf{B}\mathbf{A}^{-} \mathbf{A} = \mathbf{B}$

**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,

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

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

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

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

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

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),

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= x_1 + 2(c_1 + 4x_1) + 3(c_2 - 3x_1) = 2c_1 + 3c_2 = 5 \\ 2x_1 + x_2 + 2x_3 &= 2x_1 + (c_1 + 4x_1) + 2(c_2 - 3x_1) = c_1 + 2c_2 = 6 \end{aligned}$$

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  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$  (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  $\beta_1$  indicates a sex effect and  $\beta_2$  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  $\beta_1$  and  $\beta_2$  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

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

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  $\beta$  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  $\mathbf{b}^T \beta$  given by

$$\text{OLS}(\mathbf{b}^T \beta) = \mathbf{b}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (\text{A3.6d})$$

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

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

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

Note that  $\mathbf{X}^T \mathbf{X}$  is singular, so we cannot obtain unique estimates of all three parameters. For this design matrix, are  $\beta_3$ ,  $\beta_1 + \beta_2$ , and  $\beta_1$  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,  $\beta_3$  and  $(\beta_1 + \beta_2)$ , are estimable. However, since

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

$\beta_1$  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  $\mathbf{b}$ .

## GENERALIZED LEAST SQUARES, GLS

Under OLS, the unweighted sum of squared residuals is minimized. However, if some residuals are inherently more variable than others (have a higher variance),

less weight should be assigned to the more variable data. Correlations between residuals can also influence the weight that should be assigned to each individual, as the data are not independent. Thus, if the residual errors are heteroscedastic and/or correlated, ordinary least-squares estimates of regression parameters and standard errors of these estimates are potentially biased.

A more general approach to regression analysis expresses the covariance matrix of the vector of residuals as  $\sigma_e^2 \mathbf{R}$ , with  $\sigma(e_i, e_j) = R_{ij}\sigma_e^2$ . Lack of independence between residuals is indicated by the presence of nonzero off-diagonal elements in  $\mathbf{R}$ , while heteroscedasticity is indicated by differences in the diagonal elements of  $\mathbf{R}$ . **Generalized** (or **weighted**) **least squares** (GLS) takes these complications into account. As shown in Appendix 3, if the linear model is

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

the GLS estimate of  $\boldsymbol{\beta}$  is

$$\mathbf{b} = (\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \tag{8.34}$$

(Aitken 1935). The covariance matrix for the GLS estimates is

$$\mathbf{V}_{\mathbf{b}} = (\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X})^{-1} \sigma_e^2 \tag{8.35}$$

If residuals are independent and homoscedastic,  $\mathbf{R} = \mathbf{I}$ , and GLS estimates are the same as OLS estimates. If  $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{R}\sigma_e^2)$ , the GLS estimate of  $\boldsymbol{\beta}$  is also the maximum-likelihood estimate.

**Example 11.** A common situation requiring weighted least-squares analysis occurs when residuals are independent but heteroscedastic with  $\sigma^2(e_i) = \sigma_e^2/w_i$ , where  $w_i$  are known positive constants. For example, if each observation  $y_i$  is the mean of  $n_i$  independent observations (each with uncorrelated residuals with variance  $\sigma_e^2$ ), then  $\sigma^2(e_i) = \sigma_e^2/n_i$ , and hence  $w_i = n_i$ . Here

$$\mathbf{R} = \text{Diag}(w_1^{-1}, w_2^{-1}, \dots, w_n^{-1})$$

where  $\text{Diag}$  denotes a diagonal matrix, giving

$$\mathbf{R}^{-1} = \text{Diag}(w_1, w_2, \dots, w_n)$$

With this residual variance structure, consider the weighted least-squares estimate for the simple univariate regression model  $y = \alpha + \beta x + e$ . In GLM form,

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

Define the following weighted means and cross products,

$$w = \sum_{i=1}^n w_i, \quad \bar{x}_w = \sum_{i=1}^n \frac{w_i x_i}{w}, \quad \overline{x^2}_w = \sum_{i=1}^n \frac{w_i x_i^2}{w}$$

$$\bar{y}_w = \sum_{i=1}^n \frac{w_i y_i}{w}, \quad \overline{xy}_w = \sum_{i=1}^n \frac{w_i x_i y_i}{w}$$

With these definitions, matrix multiplication and a little simplification give

$$\mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} = w \begin{pmatrix} \bar{y}_w \\ \overline{xy}_w \end{pmatrix} \quad \text{and} \quad \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} = w \begin{pmatrix} 1 & \bar{x}_w \\ \bar{x}_w & \overline{x^2}_w \end{pmatrix}$$

Applying Equation 8.34, the GLS estimates of  $\alpha$  and  $\beta$  are

$$a = \bar{y}_w - b \bar{x}_w \tag{8.36a}$$

$$b = \frac{\overline{xy}_w - \bar{x}_w \bar{y}_w}{\overline{x^2}_w - \bar{x}_w^2} \tag{8.36b}$$

If all weights are equal ( $w_i = c$ ), these expressions reduce to the standard (OLS) least-squares estimators given by Equation 3.14. Applying Equation 8.35, the sampling variances and covariance for these estimates are

$$\sigma^2(a) = \frac{\sigma_e^2 \cdot \overline{x^2}_w}{w (\overline{x^2}_w - \bar{x}_w^2)} \tag{8.37a}$$

$$\sigma^2(b) = \frac{\sigma_e^2}{w (\overline{x^2}_w - \bar{x}_w^2)} \tag{8.37b}$$

$$\sigma(a, b) = \frac{-\sigma_e^2 \bar{x}_w}{w (\overline{x^2}_w - \bar{x}_w^2)} \tag{8.37c}$$

**Table 1.** Summary of useful results for the general linear model,  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ , under OLS and GLS assumptions for the distribution of residuals. We have discussed GLS above under the special case where  $\mathbf{V} = \sigma_e^2 \mathbf{R}$ , where  $\mathbf{R}$  is a matrix of constants.

---

Ordinary Least Squares, OLS	Generalized Least Squares, 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  $\beta$ :

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \qquad \hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

Covariance matrix for  $\hat{\beta}$ :

$$(\mathbf{X}^T \mathbf{X})^{-1} \sigma_e^2 \qquad (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$$

Predicted values,  $\hat{\mathbf{y}} = \mathbf{X} \hat{\beta}$ :

$$\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \qquad (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

Covariance matrix for predicted values,  $\hat{\mathbf{y}}$ :

$$\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma_e^2 \qquad \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T$$

$\chi^2$  goodness of fit statistic (assuming residuals are MVN):

$$\chi^2 = \frac{(\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}})}{\sigma_e^2} \qquad \chi^2 = (\mathbf{y} - \hat{\mathbf{y}})^T \mathbf{V}^{-1} (\mathbf{y} - \hat{\mathbf{y}})$$

### Derviation 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

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

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

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

where

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

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

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

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

$$\begin{aligned} \hat{\beta} &= \left( (\mathbf{X}^T \mathbf{R}^{-1/2}) (\mathbf{R}^{-1/2} \mathbf{X}) \right)^{-1} (\mathbf{X}^T \mathbf{R}^{-1/2}) (\mathbf{R}^{-1/2} \mathbf{y}) \\ &= \left( \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \end{aligned}$$

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

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

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  $\mathbf{y} - \hat{\mathbf{y}} \sim \text{MVN}(\mathbf{0}, \mathbf{V})$  and it follows from Equation A3.9 that

$$(\mathbf{y} - \hat{\mathbf{y}})^T \mathbf{V}^{-1} (\mathbf{y} - \hat{\mathbf{y}}) \sim \chi^2 \quad (\text{A3.11a})$$

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

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

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

## 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

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

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

$$\frac{\text{SSE}}{\sigma_e^2} \sim \chi_{n-p}^2 \quad (\text{A3.18})$$

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 ( $SS_{E_r} - SS_{E_f}$ ) is distributed as constant ( $\sigma_e^2$ ) times a  $\chi_{p-q}^2$ . Likewise, from Equation A3.18,  $SS_{E_f} \sim \sigma_e^2 \chi_{n-p}^2$ . Recalling the definition of the  $F$  distribution (Appendix 5), it follows that

$$\frac{(SS_{E_r} - SS_{E_f}) / (p - q)}{SS_{E_f} / (n - p)} = \left( \frac{n - p}{p - q} \right) \left( \frac{SS_{E_r}}{SS_{E_f}} - 1 \right) \tag{A3.19}$$

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  $\mu$  is  $\bar{y}$  for OLS and the weighted mean for GLS, giving  $SS_{E_r} = SS_T$ . 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

$$\left( \frac{n - p}{p - 1} \right) \left( \frac{SS_T}{SS_{E_f}} - 1 \right) = \left( \frac{n - p}{p - 1} \right) \left( \frac{r^2}{1 - r^2} \right) \tag{A3.20}$$

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.

### 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. Since these sums of squares are functions of random data, they are also themselves random values. The **observed sums of squares** are distributed about their expected values, which can be obtained by using Equation 8.22. The resulting **expected sums of squares** are functions of the unknown means and variances. By equating observed and expected sums of squares, we can often estimate the unknown parameters. This is the basic methods-of-moments procedure for estimating variance components using ANOVA (Chapter 18). Likewise, the variance of an observed sum of squares about its expected value is given by Equation 8.26 (if we are willing to accept multivariate normality assumptions).

#### Sums of Squares Expressed as Quadratic Forms

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** ( $SS_T$ ) being the sum of two components, the **error (or residual) sum of squares** ( $SS_E$ ) and the **model sum of squares** ( $SS_M$ ),

$$SS_T = SS_M + SS_E$$

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**,

$$r^2 = \frac{SS_M}{SS_T} = 1 - \frac{SS_E}{SS_T} \quad (\text{A3.15})$$

Sums of squares have different forms under OLS and GLS. Under OLS, the residuals are assumed to be independent with common variance  $\sigma_e^2$ . In this case, each observation/residual is weighted equally, with

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \bar{y}^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n y_i \right)^2$$

which can be expressed as a quadratic form of the vector of observations  $\mathbf{y}$ ,

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

where  $\mathbf{J}$  (the **unit matrix**) is  $n \times n$  and each element is 1.

Now consider the error sum of squares

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

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

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

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

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

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

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

Note that

$$SS_M = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

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

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

Likewise, the error sum of squares becomes

$$\begin{aligned} SS_E &= \hat{\mathbf{e}}^T \mathbf{R}^{-1} \hat{\mathbf{e}} \\ &= \mathbf{y}^T \left[ \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{X} \left( \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{R}^{-1} \right] \mathbf{y} \end{aligned} \quad (\text{A3.17b})$$

and the model sum of squares becomes

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

If the residuals are multivariate-normally distributed with

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

then  $SS_E/\sigma_e^2$ , the sum of squared unit normals, is  $\chi^2$ -distributed. In particular, with  $n$  observations and  $p$  estimated parameters,

$$\frac{SS_E}{\sigma_e^2} \sim \chi_{n-p}^2 \quad (8.41)$$

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