9

General Linear and Mixed Models

Version 22 December 2019

Linear models form the backbone of most estimation procedures in quantitative genetics and are extensively used throughout this book. They are generally structured such that a vector, $\mathbf{y}$, of observations of a response variable (y) is modeled as a linear combination of predictor variables observed along with y (Chapter 8). This chapter introduces some of the basic tools and key concepts underlying the use of such models. We start with models that only involve **fixed effects** (unknown constant to be estimated). Such **general linear models** (**GLM**) underlie a large number of standard statistical approaches (regression, ANOVA, etc.), all of which have estimators given by simple matrix expressions. We then introduce models with **random effects** (effect values are random draws, or **realizations**, from some underlying distribution), which requires estimating the variance of this underlying distribution in order to predict their values. Because they contain both fixed and random effects, these are called **mixed models** (**MM**), and include the GLM as a special case. Advanced linear model topics are examined in detail in Chapters 29 (BLUEs) and 30 (BLUPs), and further comments are given in Appendix 3. Appendix 9 examines experimental design in the context of linear models.

BASICS OF GENERAL LINEAR MODELS

Equation 8.1a introduced multiple regression, the extension of a univariate regression (Equation 3.12a) for predicting the value of a response variable y to using *multiple* sources of information (the predictor variables $x_1, \cdots, x_p$). In quantitative genetics, these predictors could be traits of interest, molecular marker values, or confounding **cofactors** to be removed, such as sex, block, or age effects. Some or all of these predictors could be **indicator variables**, with values of 0 or 1 indicating whether an observation belongs in a particular category or grouping of interest. Models containing only indicator variables are often 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 are special cases of the **general linear model** (**GLM**), wherein each observation (y) is assumed to be a linear function of m observed and/or indicator variables plus a residual error,

$$y_i = \mu + \sum_{k=1}^{m} \beta_k \, x_{ik} + e_i \tag{9.1}$$

where $x_{i1}, \cdots, x_{im}$ are the values of the m predictor variables for the *i*th individual.

Linear Models in Matrix Form

For a vector of n observations, Equation 9.1 can be compactly written in matrix form as

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times p}\,\boldsymbol{\beta}_{p\times 1} + \mathbf{e}_{n\times 1} \tag{9.2}$$

where $\mathbf{y}$ is the $(n \times 1)$ vector of observed y values, the **design** or **incidence matrix** $\mathbf{X}$ is $n \times p$, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of parameters (usually called **factors** or **effects**) to be estimated, and $\mathbf{e}$

is the vector of **residual errors** (or, simply, **residuals**). Here $\mathbf{y}$ and $\mathbf{X}$ are known (the data), while $\boldsymbol{\beta}$ is estimated, and then used to infer the residuals. For a given value of $\boldsymbol{\beta}$,

$$\widehat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} \quad \text{and} \quad \widehat{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}} \tag{9.3}$$

where $\widehat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}$ is the $(n \times 1)$ vector of predicted values for a given model.

Before examining how to solve a general linear model (i.e., estimation of $\boldsymbol{\beta}$), we first present a number of examples to give the reader a feel of how various linear models are expressed in matrix form. This compact representation follows from the definition of matrix multiplication (Equation 8.6a).

Example 9.1. Consider a multiple regression using k predictor variables,

$$y_i = \alpha + \sum_{j=1}^{k} \beta_j \, x_{ij} + e_i \tag{9.4}$$

where x_{ij} is the value of the jth predictor variable in observation i. We can think of the ith observation as having two components: the value y_i of a response variable to be predicted, and the vector $\mathbf{x}_i^T = (1, x_{i1}, \cdots, x_{ik})$ of predictor values for that observation. For n observations, this model can be expressed in GLM matrix form as

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \text{and} \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

Note from matrix multiplication that $y_i = \mathbf{x}_i^T\boldsymbol{\beta} + e_i$ recovers Equation 9.4. The ith row of $\mathbf{X}$ ($\mathbf{x}_i^T$) corresponds to the values of the predictor variables from observation i, while the jth column of $\mathbf{X}$ corresponds to the data on parameter j in the experiment. The interpretation of β_i is the change expected in y from a unit change in the predictor variable x_i, while holding all other predictor variables constant.

As a specific example, consider a regression with three predictor variables,

$$y_i = \mu + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_1 x_{i3} + e_i$$

The predicted value becomes

$$\widehat{y}_i = \mu + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_1 x_{i3}$$

with $e_i = y_i - \widehat{y}_i$ being the difference between the observed and predicted values. In matrix form, this model becomes

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{pmatrix} \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Example 9.2. Consider a half-sib design wherein each of m sires (males) are mated at random to a number of unrelated dams (females) and a single offspring is measured from each cross. The simplest model for this design (i.e., ignoring any other potential cofactors) is

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

where y_{ij} is the phenotype of the *j*th offspring from sire *i*, μ 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 identical to the standard regression model by using *m* **indicator variables** to classify the sires of the offspring. The resulting linear model becomes

$$y_{ij} = \mu + \sum_{k=1}^{m} s_k\, x_{ik} + e_{ij}, \quad \text{where} \quad x_{ik} = \begin{cases} 1 & \text{if sire } k = i \\ 0 & \text{otherwise} \end{cases}$$

Suppose that three different sires used in the above half-sib design have two, one, and three offspring, respectively. In matrix form, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with

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

This simple example highlights two issues. First, the model is **overparameterized**, as one could arbitrary choose a value for μ and adjust each of the sire effect according. We can see this from the column structure of $\mathbf{X}$, as its first column is simply the sum of columns two through four. Hence, $\mathbf{X}$ has only has three independent columns (it has rank three) and can only uniquely estimate three parameter combinations. The rank of $\mathbf{X}$ impacts the estimatibility of model parameters, which we will examine more formally, both below and in Appnedix 3. We can easily deal with this issue with the modified model

$$y_{ij} = t_i + e_{ij}$$

where $t_i = \mu + s_i$ is the mean value for sire i, with $\mathbf{X}$ and $\boldsymbol{\beta}$ now becoming

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

The second issue is more subtle, but equally important. Our above treatment assumes that the sire values are **fixed effects** (unknown constants), but in reality they are best treated as **random effects** (realizations of random draws from some unknown distribution). As detailed below, this shift in focus from fixed to random introduces issues in interpretation, but also allows for more efficient use of the data. For example, if the sires are related, then information on a related sire (say sire 3) further informs us as to the value for sire 1. Under a fixed-effect framework, the only information we can use to estimate the sire 1 effect are *direct observations* on that sire. In this example, those are only the first two observations (only those observations with a 1 in the first column of $\mathbf{X}$, corresponding to data on sire 1). Under a random-effects framework we can also ***burrow information from correlated (indirect) observations***. Hence, if sires 1 and 3 are related, information from the last three observations (direct observations on sire 3) provides additional information on sire 1 that is not accessed in a fixed-effect analysis. We examine fixed versus random effects in detail shortly, with Example 9.10 reframing this example in a random-effects setting.

Example 9.3. Consider a **polynomial regression**,

$$y_i = a + \beta_1 x_i + \beta_2 x_i^2 + \cdots \beta_k x_i^k + e_i \tag{9.5}$$

Although the data enter as nonlinear functions, this is still a GLM, as the *linear* in GLM refers to linear with respect to the *parameters being estimated*, not the data. In the case of a polynomial regression, the resulting GLM matrices become

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

More generally, any regression of the form

$$y_i = \alpha + \beta_1 f(x_i) + \beta_2 g(x_i) + \cdots + e \tag{9.6}$$

can be placed in GLM form. Conversely,

$$y_i = \alpha + \exp(-\beta x_i)$$

is *not* a linear model, as it is a nonlinear function of the parameter β. In some cases, a transformation can recover a linear model. For example, the model

$$y_i = \alpha \exp(-\beta x_i)$$

can be written in linear model form as

$$\log(y_i) = \log(\alpha) - \beta x_i$$

Example 9.4. A common addition to many GLMs are **interaction terms**, such as the model

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i2} x_{i1} + e_i \tag{9.7}$$

Here

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

The interpretation of an interaction term is that the effect of some of the predictor variables on y are *context-specific*, and hence change as the background values of other predictor variables change. In this example, when x_1 held constant, a unit change in x_2 changes y by $\beta_2 + \beta_3 x_1$ (i.e., the slope of x_2 depends on the current value of x_1). Likewise, a unit change in x_1 changes y by $\beta_1 + \beta_3 x_2$.

Example 9.5. One major strength of GLMs is their ***incredible flexibility***, as one can mix and match all of the above modifications (and them some!) to design a model motivated by biological intuition, rather than statistical convenience. As an example, suppose you want a quadratic regression forced through the origin, with the slope of the quadratic term varying over the sexes. The resulting GLM can be written as

$$y_i = \beta_1 x_i + \beta_2 x_i^2 + \beta_3 s_i x_i^2 + e_i$$

Here s_i is an indicator (0/1) variable for the sex (0 = male, 1 = female). This yields a male quadratic slope of β_2 and a female quadratic slope of $\beta_2 + \beta_3$. The standard linear model hypothesis-testing framework (Appendix 3) can be used to test whether β_3 is significantly different from zero.

Solving Linear Models: Ordinary Least Squares (OLS)

Estimates of the vector $\boldsymbol{\beta}$ of unknown parameters for a general linear model are usually obtained by the method of least-squares (Chapter 3). This approach uses the observational data given by $\mathbf{y}$ and $\mathbf{X}$ and makes assumptions about the covariance structure of the vector of residual errors, $\mathbf{e}$, leading to either **ordinary least squares** (**OLS**) or **general least squares** (**GLS**) solutions. Both approaches estimate $\boldsymbol{\beta}$ by minimizing a sum of squared residuals, with OLS minimizing an unweighted sum, and GLS a weighted sum, as we now detail.

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

$$\widehat{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b} \tag{9.8a}$$

The OLS estimate of $\boldsymbol{\beta}$ is the $\mathbf{b}$ vector that minimizes the *unweighted* residual sum of squares,

$$\sum_{i=1}^{n} \widehat{e}_i^2 = \widehat{\mathbf{e}}^T\widehat{\mathbf{e}} = (\mathbf{y} - \mathbf{X}\mathbf{b})^T(\mathbf{y} - \mathbf{X}\mathbf{b}) \tag{9.8b}$$

Taking derivatives, Example A3.6 shows that the OLS estimate satisfies

$$\text{BLUE}_{OLS}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{9.9a}$$

with the OLS estimator of β_i given by the ith element of this column vector. Equation 9.9a is called the **BLUE** (or, more precisely, the **OLS BLUE**), for **best linear unbiased estimator**, of $\boldsymbol{\beta}$.

As an aside

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T \tag{9.9b}$$

is referred to as the **hat matrix** (Hoaglin and Welsch 1978), because, for an OLS estimate, the vector of predicted y values is given by

$$\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{H}\mathbf{y} \tag{9.9c}$$

$\mathbf{H}$ maps the observed $\mathbf{y}$ values onto their predicted values $\widehat{\mathbf{y}}$. Likewise, define

$$\mathbf{M} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T \tag{9.9d}$$

as the **absorption matrix** for the fixed effects. Here

$$\mathbf{M}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{e} \tag{9.9e}$$

By minimizing the unweighted sum, OLS assumes that all residuals contain the same amount of information. This implies that the residuals are homoscedastic, with $\sigma^2(e_i) = \sigma_e^2$ for all i. If some of the residuals have smaller variances, then the corresponding predicted values of y are more precise, and hence their associated residuals should be weighed more heavily that residuals with larger variances. Using an unweighted sum further assumes that all residuals are uncorrelated, $\sigma(e_i, e_j) = 0$ for $i \neq j$, as the presence of correlated residuals would change the residual weighting scheme. The OLS assumptions about the residual structure is compactly denoted by $\mathbf{e} \sim (\mathbf{0}, \sigma_e^2\,\mathbf{I})$, namely, that the vector of residuals has a mean vector of zero, and a variance-covariance matrix of

$$\text{Var}(\mathbf{e}) = \sigma_e^2\,\mathbf{I} = \begin{pmatrix} \sigma_e^2 & 0 & \cdots & 0 \\ 0 & \sigma_e^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_e^2 \end{pmatrix} \tag{9.10}$$

Under the OLS assumption, the covariance matrix of the elements of vector of BLUE estimates is

$$\mathbf{V_b} = \sigma_e^2 \, (\mathbf{X}^T\mathbf{X})^{-1} \tag{9.11a}$$

Hence, the sampling variance for the estimate of β_i is given by the ith diagonal element of the matrix $\mathbf{V_b}$, while the covariance of this estimator with the BLUE of β_j is the ijth element of $\mathbf{V_b}$. In applying Equation 9.11a, the residual variance σ_e^2 is usually replaced by the estimated variance,

$$\widehat{\sigma}_e^2 = \frac{1}{n-p}\sum_{i=1}^{n}\widehat{e}_i^2 = \frac{(\mathbf{y}-\mathbf{Xb})^T(\mathbf{y}-\mathbf{Xb})}{n-p} \tag{9.11b}$$

where $\mathbf{b}$ is the BLUE estimate (Equation 9.9a) and p parameters are estimated (p is replaced by the rank of $\mathbf{X}$, when rank[$\mathbf{X}$] $< p$).

It is important to stress that the *only* assumption required to obtain Equations 9.9a and 9.11a is that the covariance matrix for the residuals is $\sigma_e^2\,\mathbf{I}$. While it is often assumed that the residuals must also be normally distributed, this is *not* required for Equations 9.9a and 9.11a to hold. The MVN assumption, however, is required for hypothesis testing and construction of confidence intervals (Appendix 3). If the residuals follow a multivariate normal distribution with $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \sigma_e^2\,\mathbf{I})$, then the OLS estimate is also the maximum-likelihood estimate (Appendix 4).

Finally, we note that the structure of $\mathbf{X}$ informs one as to whether the data contains enough information to uniquely estimate a particular parameter. A parameter is said to be **estimable** if the model returns a unique estimate of its value. If $(\mathbf{X}^T\mathbf{X})^{-1}$ exists, then every parameter in a GLM is estimable. What happens when $\mathbf{X}^T\mathbf{X}$ is singular? As discussed in Appendix 3, Equations 9.9a and 9.11a still hold when a **generalized inverse** (or **g-inverse**), denoted $(\mathbf{X}^T\mathbf{X})^-$, is used. In this case, not all of the parameters have unique estimates, rather some can only be uniquely estimated as linear combinations (or **contrasts**). For example, a three-parameter model might return an estimate of $\beta_1 = 3$, but only be able to specify that (say) $\beta_2 - 3\beta_3 = 2$. Hence, only two unique *parameter combinations* can be estimated from the data.

Lack of estimability often implies a poor experimental design, but can also arise through loss of data from an otherwise well-planned design. Suppose one is examining the effects of height and sex. If the data (from either poor initial design or through data loss) consists of only tall males and short females, one cannot separate these effects from each other. Rather, we can only contrast tall males with short females, with the height and sex effects being fully confounded. The number of unique combinations of fixed effects that can be estimated for a given model is given by the rank of $\mathbf{X}$, namely the number of independent columns (columns that cannot be expressed as linear combinations of the other columns; Example 9.2). With n observations and $p < n$ unknowns, $\mathbf{X}$ is an $n \times p$ matrix, so that $\mathbf{X}^T\mathbf{X}$ is $p \times p$. $\mathbf{X}$ is of full rank when its rank is p, and in such cases the inverse of $\mathbf{X}^T\mathbf{X}$ exists. When this inverse does not exists (so that Rank[$\mathbf{X}$] $< p$), the rank of $\mathbf{X}$ is given by the number of nonzero eigenvalues of $\mathbf{X}^T\mathbf{X}$. When the rank of $\mathbf{X}$ is less than p, a g-inverse must be used. Appendix 3 examines these issues in more detail.

Example 9.6. 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

$$y_i = \beta\, x_i + e_i \tag{9.12a}$$

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

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

Assuming the covariance matrix of $\mathbf{e}$ is $\sigma_e^2\,\mathbf{I}$, then Equation 9.9a gives the OLS estimate of β as

$$\widehat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \frac{\sum x_i\,y_i}{\sum x_i^2} \tag{9.12b}$$

while Equations 9.11a and 9.11b give its sample variance as

$$\sigma^2(\,\widehat{\beta}\,) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\sigma_e^2 = \frac{\sigma_e^2}{\sum x_i^2} \simeq \frac{\sum_{i=1}^{n}(y_i - \widehat{\beta}x_i)^2}{(n-1)\sum x_i^2} \tag{9.12c}$$

Note that this estimate of β differs from the standard univariate regression slope (Equation 3.14b) where the intercept value is not assumed to be equal to zero.

Example 9.7. Recall from Equation 8.4b 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 $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. For the ith individual, we observe y_i and the values of p predictor variables, $z_{i1}, \cdots, z_{ip}$. Because a regression satisfies $\bar{y} = \alpha + \beta_1\bar{z}_1 + \cdots + \beta_p\bar{z}_p$, subtracting the mean from each observation removes the intercept, with

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

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

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

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

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

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

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

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

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

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 8.4b does indeed give the OLS estimates of the partial regression coefficients.

Solving Linear Models: Generalized Least Squares (GLS)

Under OLS, the unweighted sum of squared residuals is minimized, which assumes homoscedastic and uncorrelated residuals. When 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, and minimizes the weight sum of squares given by $\mathbf{e}^T\mathbf{R}^{-1}\mathbf{e}$. As shown in Appendix 3, this follows as the transformation $\mathbf{e}_* = \mathbf{R}^{-1/2}\mathbf{e}$ changes the original vector of residuals into a new vector $\mathbf{e}_*$ with homoscedastic and uncorrelated residuals, with the GLS solution following from the minimization of the sum $\mathbf{e}_*^T\mathbf{e}_*$. As a result, for the linear model

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

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

$$\text{BLUE}_{GLS}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \tag{9.13a}$$

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

$$\mathbf{V_b} = \sigma_e^2\left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1} \tag{9.13b}$$

The estimated residual variance is also modified from Equation 9.11b, with

$$\widehat{\sigma}_e^2 = \frac{(\mathbf{y}-\mathbf{X}\mathbf{b})^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})}{n - \text{rank}(\mathbf{X})} \tag{9.13c}$$

If residuals are independent and homoscedastic, $\mathbf{R} = \mathbf{I}$, and GLS estimates are the same as OLS estimates. If $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \sigma_e^2\,\mathbf{R})$, the GLS estimate of $\boldsymbol{\beta}$ is also the maximum-likelihood estimate. In our treatment of GLS, we will occasionally denote the residual covariance structure by $\mathbf{V} = \sigma_e^2\,\mathbf{R}$.

Example 9.8. 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 σ_e^2), then $\sigma^2(e_i) = \sigma_e^2/n_i$, and hence $w_i = n_i$. Here

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

where Diag denotes a diagonal matrix, giving

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

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,

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

Define the following weighted means and cross products,

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

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

With these definitions, matrix multiplication and a little simplification yields

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

Applying Equation 9.13a, the GLS estimates of α and β are

$$a = \overline{y}_w - b\overline{x}_w \tag{9.14a}$$

$$b = \frac{\overline{xy}_w - \overline{x}_w\,\overline{y}_w}{\overline{x^2}_w - \overline{x}_w^2} \tag{9.14b}$$

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 9.13b, the sampling variances and covariance for these estimates become

$$\sigma^2(a) = \frac{\sigma_e^2 \cdot \overline{x^2}_w}{w\,(\,\overline{x^2}_w - \overline{x}_w^2\,)} \tag{9.15a}$$

$$\sigma^2(b) = \frac{\sigma_e^2}{w\,(\,\overline{x^2}_w - \overline{x}_w^2\,)} \tag{9.15b}$$

$$\sigma(a,b) = \frac{-\sigma_e^2\,\overline{x}_w}{w\,(\,\overline{x^2}_w - \overline{x}_w^2\,)} \tag{9.15c}$$

Model Assessment

Fitting a model is only the first step, and should be followed by considerable post-construction analysis. A few key ideas. First, just because a model is *statistically significant* (it has a significantly better fit than a model with no effects beyond a common mean; $y_i = \mu + e_i$) does not mean that it is *biologically significant*. A good indicator of the later is the model r^2 value (Equation A3.15), namely, how much of the variation in the response variable (y) is accounted for by the model. Consider two different models: Model 1 with $p = 10^{-9}$ and $r^2 = 0.01$, while Model 2 has $p = 10^{-4}$ and $r^2 = 0.3$. From a statistical standpoint, Model 1 might seem superior, but it only accounts for 1% of the variation in y, while Model 2 accounts for 30% and thus is biologically more significant than Model 1.

Second, all of the information on how well the model performs is contained in the residuals. As mentioned, normality is *not* required for OLS or GLS estimates or their standard errors. It is, however, required for hypothesis testing and the construction of confidence intervals (Appendix A3). Hence, the vector of residuals should be tested for normality, at a minimum by plotting their histogram, and more formally by using standard normality tests (Chapter 2). It should be noted that a histogram of the y values is usually *expected* to be nonnormal, as the data likely consists of a number of groups with different means. Even if

y values are normally distributed in each of these groups, the distribution of y values in the full sample is a mixture of normals, and hence is likely to show multiple modes, skew, and kurtosis (Chapter 15). In contrast, the distribution of residuals reflect errors once we have accounted for the data consisting of different groups with different means. If one finds that the residuals are not normal, transformation of y to a different scale (Chapter 13), such as constructing a GLM using $\log(y)$ in place of y, can often normalize the resulting residuals. More generally, one can used generalized linear models (Chapter 14) to handle settings where the residuals are *expected* to be non-normally, such as with binary data.

Finally, **residual plots** of e versus y, or e versus $\widehat{y}$, can be highly informative, as can a plot of $\widehat{y}$ versus y. While (by construction) the expected value of e is zero under least-squares, there may still be trends in e despite this restriction. For example, if the spread of residual values increases with y or $\widehat{y}$, this indicates heteroscedasticity (and hence GLS must be used instead of OLS). Modeling using $\log(y)$ instead of y can often resolve this issue. More generally, the distribution of e should be independent over y values if one has an appropriate model. Given that one observes a positive residual, an adjacent residual (when plotting e vs. y or $\widehat{y}$) is equal likely to be positive or negative under a valid model. Departures from this independence can arise if an incorrect model is fitted. As an example, suppose that the true model is a quadratic (with a maximum in the middle of the data range), but we assumed a linear effect. The latter model would likely still account for some of the variation in y. However, residual plot would show blocks (**runs**) of residuals all having the same sign, for example, the first $1/4$ are negative, the middle $1/2$ are positive, and the last $1/4$ are negative. This would still give an expected residual value of zero, but adjacent residuals are correlated due to fitting an incorrect model. A simple check for this is the **Wald–Wolfowitz runs test**. Here the test statistic, R_n, for the number of runs (changes in the signs of adjacent residuals) in a sample of size n is approximately normally distributed with mean $n/2 + 1$ and variance $n/4$.

THE GENERAL MIXED MODEL

The **mixed model** (**MM**) builds on the general linear model by adding one (or more) vectors of random effects,

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

where $\mathbf{Z}$ (akin to $\mathbf{X}$) is a known matrix and $\mathbf{u}$ is the additional vector of random effects. As with the residuals of a general linear model, one must ***specify the covariance structure*** of $\mathbf{u}$ in order to define the model. Before proceeding further with the analysis of the MM—which is the workhorse framework for much of modern quantitative genetics—we first consider the distinction between fixed versus random effects.

Fixed Versus Random Effects

Unknown parameters in statistical models are typically classified as either **fixed** or **random** (Figure 9.1). The former are unknown constants to be estimated, whereas the latter are realizations drawn from some underlying distribution. While this classification is often treated rather rigidly, the distinction between the two is actually somewhat fluid, and depends on often subtle differences in the assumptions underlying a model. Indeed, for the same data set and essentially the same problem, one investigator might feel justified using a fixed-effect interpretation, while another might choose random effects.

One common rule is that if one regards the data as a sample from a much large population of interest, then effects are treated as random. If our interest is only in the values in the sample, then often (but not always; see Example 9.9) they are treated as fixed. Likewise, if one cannot go back after the fact to resample the same **levels** of a factor, it is best treated as random. For example, one can go back and sample new males and females, or resample from a specific growing location. However, one cannot resample from a specific past year.

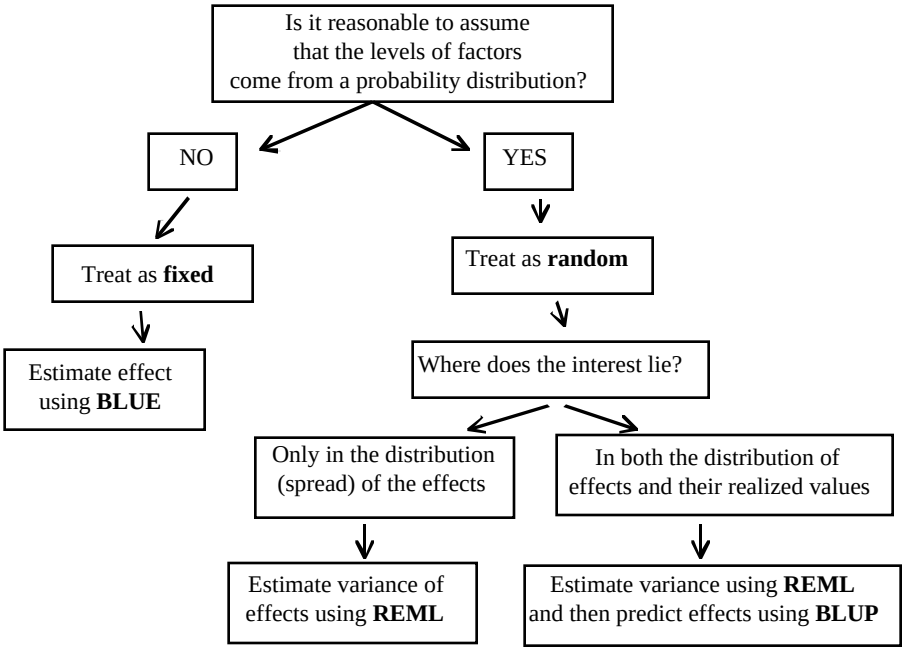

Figure 9.1. A flow chart for considering the roles of fixed and random effects in an analysis. (From Hans-Peter Piepho, reprinted with permission.)

Hence, one could treat the location as fixed, but the year as random (and hence a year by location interaction is also random).

Finally, factors are often treated as random for convenience (Example 9.9). One setting for this is when a model contains a large number of **nuisance parameters**: parameters that must be estimated, but are not really of any concern to the investigator. Each estimated fixed effect uses a degree of freedom. In contrast, random effects represent realizations (draws) from a particular unknown distribution, typically assumed to be a normal with mean zero (the fixed effects absorb any nonzero mean effect) and variance σ^2. Here, the only parameter to be estimated is this variance, *no matter how many realizations are drawn* from this distribution. Thus, nuisance parameters are often treated as random to save degrees of freedom. Finally, one can always move to a Bayesian framework (Appendices 7 and 8), wherein all parameters are treated as random.

Example 9.9. Suppose that one has a completely randomized design (Appendix 9) of 10 blocks, each of which consists of three replicate plots for each of five different inbred lines, randomly assigned within the block (for 150 observed plot values). The simplest linear model is

$$y_{ijk} = L_i + B_j + e_{ijk} \tag{9.16}$$

with y_{ijk} the value of the kth plot of line i in block j, and we assume the absence of line $\times$ block interactions. Treating all factors as fixed, there are five line effects and 10 block effects, or 15 parameters (degrees of freedom) that must be estimated (the column vector $\boldsymbol{\beta}$ has 15 terms). Assuming OLS, the covariance structure becomes $\mathbf{e} \sim (0, \sigma_e^2\, \mathbf{I})$.

However, we usually do not care about the block effects beyond removing any bias they may introduce in the estimate of the L_i. If we could reasonably assume that blocks have independent effects (i.e., blocks that are closer to each other are no more likely to be similar than comparisons among more distant blocks), then we can treat these as random, with the value (or realization) for a particular block being drawn from some distribution with unknown variance, σ_b^2. The consequence of treating them as random is that instead of using 10 degrees

of freedom for block effects, we *only use one*, that for estimating σ_b^2. Similarly, even with 1000 blocks, we would still only use one degree of freedom under the random-effect model (assuming that all the blocks are uncorrelated).

Under this interpretation, $\boldsymbol{\beta}$ has five terms (one for each line, with $\mathbf{X}$ being 150×5), while the column vector $\mathbf{u}$ of random effects has ten terms (one for each block, with $\mathbf{Z}$ being 150×10). The model is still given by Equation 9.16, but we have to further specify the covariance structure for the random block effects (along with the residuals). Assuming independent block effects and OLS residuals,

$$\mathbf{b} \sim (0, \sigma_b^2\,\mathbf{I}), \qquad \mathbf{e} \sim (0, \sigma_e^2\,\mathbf{I}) \tag{9.17a}$$

If one assumes correlation among adjacent blocks, then $\mathbf{I}$ is replaced by a nondiagonal matrix whose off-diagonal elements indicate the amount of correlation between each combination of blocks. This requires adding (at least) one additional parameter to estimate, with the random effects now using (at least) two degrees of freedom.

The model (covariance structure) given by Equation 9.17a still treats the genotypes as fixed, which implies that we only care about the values of the particular lines in our sample. We have no interest in the larger population from which they were drawn. Conversely, we indeed be interested in the population variance, σ_G^2, of line values, treating our five lines as random draws from this distribution. This is now a random-effects interpretation, which introduces two important features. First, we expect that some lines may be more similar than others due to more recent common ancestry (which can be estimated using molecular markers; Chapter 16). Hence, the covariance matrix need not be diagonal, but rather of the form $\sigma_G^2\mathbf{A}$, where the **relationship matrix** $\mathbf{A}$ is symmetric. The resulting covariance structure becomes

$$\mathbf{L} \sim (0, \sigma_G^2\,\mathbf{A}), \quad \mathbf{b} \sim (0, \sigma_b^2\,\mathbf{I}), \qquad \mathbf{e} \sim (0, \sigma_e^2\,\mathbf{I})$$

This is more compactly written as

$$\begin{pmatrix} \mathbf{L} \\ \mathbf{b} \\ \mathbf{e} \end{pmatrix} \sim \begin{pmatrix} \sigma_G^2\,\mathbf{A} & 0 & 0 \\ 0 & \sigma_b^2\,\mathbf{I} & 0 \\ 0 & 0 & \sigma_e^2\,\mathbf{I} \end{pmatrix} \tag{9.17b}$$

Second, a critical feature about $\mathbf{A}$ is that random-effects models *burrow information from correlated observations*. When treating lines as fixed effects, the only information used to estimate a particular line value (say L_3) are those direct observations on line 3 (30 in this example). Under a random-effect model, direct observations on line 3 are used, as well as observations on *all* other lines that are correlated with line 3 (those lines for which the elements A_{3i} and A_{i3} of $\mathbf{A}$ are nonzero, i.e., those lines related to line 3). Hence, even if we have no interest in the nature of the population from which the lines were drawn, treating them as random allows us to access this burrowing of information feature.

The General Mixed Model

Consider a column vector $\mathbf{y}$ containing the phenotypic values for a trait measured in n individuals. We assume that these observations are described adequately by a linear model with a $p \times 1$ vector of fixed effects ($\boldsymbol{\beta}$) and a $q \times 1$ vector of random effects ($\mathbf{u}$). The first element of the vector $\boldsymbol{\beta}$ is typically the population mean, and other factors included may be gender, location, year of birth, experimental treatment, and so on. The elements of the vector $\mathbf{u}$ of random effects are often genetic effects such as additive genetic (breeding) values. In matrix form,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{9.18}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are, respectively, $n \times p$ and $n \times q$ **incidence matrices** ($\mathbf{X}$ is also called the **design matrix**), and $\mathbf{e}$ is the $n \times 1$ column vector of residual deviations assumed to be distributed independently of the random genetic effects. Often, all of the elements of $\mathbf{Z}$ are 0 or 1, depending upon whether the relevant effect for a given observation contributes to the

value of the response variable, y. Because this model jointly accounts for fixed and random effects, it is generally referred to as a **mixed model** (Eisenhart 1947).

Now consider the means and variances of the component vectors of the mixed model. By definition, random effects have mean zero. Hence, $E(\mathbf{u}) = E(\mathbf{e}) = \mathbf{0}$, implying that $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$. Denote the $(n \times n)$ covariance matrix for the vector $\mathbf{e}$ of residual errors by $\mathbf{R}$ and the $(q \times q)$ covariance matrix for the vector $\mathbf{u}$ of random genetic effects by $\mathbf{G}$. Excluding the difference among individuals due to fixed effects, from Equation 8.21b and the assumption that $\mathbf{u}$ and $\mathbf{e}$ are uncorrelated, the covariance matrix for the vector of observations $\mathbf{y}$ is

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} \tag{9.19}$$

The first term accounts for the contribution from the vector $\mathbf{u}$ of random effects, while the second accounts for the variance due to residual effects. We will generally assume that residual errors have constant variance and are uncorrelated, so that $\mathbf{R}$ is a diagonal matrix, with $\mathbf{R} = \sigma_E^2\,\mathbf{I}$. Hence, $\mathbf{X}\boldsymbol{\beta}$ gives the vector of mean values for the observations, while the spread about this mean value is influenced by the random effects, $\mathbf{u}$ and $\mathbf{e}$. The ith diagonal element of $\mathbf{V}$ (V_{ii}) gives the variance associated with observation i, while V_{ij} is the covariance between observations i and j.

We are now in a position to contrast the mixed model and the general linear model. Under the general linear model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}^* \qquad \text{where} \quad \mathbf{e}^* \sim (\mathbf{0}, \mathbf{V}) \quad \text{implying} \;\; \mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V}) \tag{9.20a}$$

where the notation $\sim (a, b)$ means that the random variable has mean a and variance b. On the other hand, the mixed model partitions the vector of residual effects into two components, with $\mathbf{e}^* = \mathbf{Z}\mathbf{u} + \mathbf{e}$, giving

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

$$\text{implying} \;\; \mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V}) = (\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}) \tag{9.20b}$$

When analyzed in the appropriate way, both formulations yield the same estimate of the vector of fixed effects $\boldsymbol{\beta}$, while the mixed-model formulation further allows estimates of the vector of random effects $\mathbf{u}$.

Comparing Equations 9.20a with 9.20b shows one important use of adding random effects. In Equation 9.20a, the vector of residuals $\mathbf{e}^*$ is of GLS form, while by adding an appropriate random effect in Equation 9.20b, the vector of residuals $\mathbf{e}$ is now in OLS form. Shared features (such as some individuals sharing the same family) generate correlations among residuals, but these can be removed by adding the appropriate vector (or vectors) of random effects to the model (Chapter 29).

For the mixed model, we observe $\mathbf{y}$, $\mathbf{X}$, and $\mathbf{Z}$, while $\boldsymbol{\beta}$, $\mathbf{u}$, $\mathbf{R}$, and $\mathbf{G}$ are generally unknown. Thus, mixed-model analysis involves two complementary estimation issues: (1) estimation of the vector of fixed effects ($\boldsymbol{\beta}$) and prediction of the vector of random effects ($\mathbf{u}$), and (2) estimation of the covariance matrices $\mathbf{G}$ and $\mathbf{R}$. These covariance matrices are generally assumed to be functions of a few unknown variance components. Namely, they are usually of the form $\sigma^2\,\mathbf{B}$, where σ^2 is an unknown variance to be estimated and $\mathbf{B}$ is a matrix of known constants. More generally, they can be of the form $\sum_i \sigma_i^2\,\mathbf{B}_i$, where the sum is usually over just a few terms. For the remainder of the chapter, we consider estimators of $\boldsymbol{\beta}$ and $\mathbf{u}$ under the assumption that the variance components associated with $\mathbf{G}$ and $\mathbf{R}$ are known. Applications of mixed models for estimating breeding values and for genomic selection and prediction are examined in detail in Chapter 29, while estimation of the variance components is examined in detail in Chapter 30.

Example 9.10. As an expansion of Example 9.2, suppose that three sires are chosen at random from a population, and each crossed to multiple randomly chosen (and unrelated) dam, where

each mating yields a single offspring. Two offspring from each sire are evaluated, some in environment 1 and some in environment 2. Let y_{ijk} denote the phenotypic value of the kth offspring of sire i in environment j. The model is then

$$y_{ijk} = \beta_j + u_i + e_{ijk}$$

This model has three random effects (u_1, u_2, u_3), which measure the contribution from each sire, and two fixed effects (β_1, β_2), which describe the influence of the two environments. The model assumes an absence of sire $\times$ environment interaction. As noted above, a total of six offspring were measured. One offspring of sire 1 was assigned to environment 1 and had phenotypic value $y_{1,1,1} = 9$, while the second offspring was assigned to environment 2 and had phenotypic value $y_{1,2,1} = 12$. The two offspring of sire 2 were both assigned to environment 1 and had values of $y_{2,1,1} = 11$ and $y_{2,1,2} = 6$. One offspring of sire 3 was assigned to environment 1 and had phenotypic value $y_{3,1,1} = 7$, while the second offspring was assigned to environment 2 and had phenotypic value $y_{3,2,1} = 14$. The resulting vector of observations can be written as

$$\mathbf{y} = \begin{pmatrix} y_{1,1,1} \\ y_{1,2,1} \\ y_{2,1,1} \\ y_{2,1,2} \\ y_{3,1,1} \\ y_{3,2,1} \end{pmatrix} = \begin{pmatrix} 9 \\ 12 \\ 11 \\ 6 \\ 7 \\ 14 \end{pmatrix}$$

giving the mixed model as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, where the incidence matrices for fixed and random effects and the vectors of these effects are

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \qquad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$$

Note that this model is not yet properly specified, as no covariance structure of $\mathbf{u}$ or $\mathbf{e}$ has been proposed.

Estimating Fixed Effects and Predicting Random Effects

As highlighted throughout this book, one primary goal of a quantitative-genetic analysis is to estimate variance components. However, there are also numerous situations in which inferences about fixed effects (such as the effect of a particular environment or year) and/or random effects (such as the breeding value of a particular individual) are the central motivation. As we will see, marker effects can be treated as either fixed or random, depending upon the circumstances (Chapters 19, 29, and 30). Inferences about fixed effects have come to be called **estimates,** whereas those that concern random effects are known as **predictions.** The most widely used procedures are BLUE and BLUP, referring, respectively, to **best linear unbiased estimator** and **best linear unbiased predictor.** They are *best* in the sense that they minimize the sampling variance, *linear* in the sense that they are linear functions of the observed phenotypes $\mathbf{y}$, and *unbiased* in the sense that $\mathrm{E}[\,\mathrm{BLUE}(\boldsymbol{\beta})\,] = \boldsymbol{\beta}$ and $\mathrm{E}[\,\mathrm{BLUP}(\mathbf{u})\,] = \mathbf{u}$. Although the method of predicting random effects using BLUP methodology was first discussed by Henderson (1949, 1950), the expression "best linear unbiased predictor" was apparently first used by Goldberger (1962), with the acronym BLUP due to Henderson (1973). BLUP is the method of choice for estimating the breeding values of individuals from field records of large and complex pedigrees, and underpins many marker-based genomic prediction and genomic selection schemes (Chapter 29).

For the mixed model given by Equation 9.18, the BLUE of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \tag{9.21}$$

with $\mathbf{V}$ as given by Equation 9.19. Notice that this is just the generalized least-squares (GLS) estimator (Equation 9.13a).

We obtain the BLUP for the vector of random effects $\mathbf{u}$ from its regression on $\mathbf{y}$. This follows from Equation 8.29a under the assumption that the joint distribution of $\mathbf{u}$ and $\mathbf{y}$ is multivariate normal. Starting with

$$\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{G}), \quad \mathbf{y} \sim \text{MVN}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

and noting (from Equation 8.21a) that the covariance between $\mathbf{u}$ and $\mathbf{y}$ is given by

$$\text{Cov}(\mathbf{u}, \mathbf{y}) = \text{Cov}(\mathbf{u}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}) = \text{Cov}(\mathbf{u}, \mathbf{u})\,\mathbf{Z}^T = \mathbf{G}\mathbf{Z}^T \tag{9.22a}$$

The joint density of $\mathbf{u}$ and $\mathbf{y}$ thus becomes

$$\begin{pmatrix}\mathbf{u}\\ \mathbf{y}\end{pmatrix} \sim \text{MVN}\left[\begin{pmatrix}\mathbf{0}\\ \mathbf{X}\boldsymbol{\beta}\end{pmatrix}, \begin{pmatrix}\mathbf{G} & \mathbf{G}\mathbf{Z}^T\\ \mathbf{Z}\mathbf{G} & \mathbf{V}\end{pmatrix}\right] \tag{9.22b}$$

In the notation of Equation 8.29a,

$$\mathbf{x_1} = \mathbf{u}, \quad \mathbf{x_2} = \mathbf{y}, \quad \boldsymbol{\mu_1} = \mathbf{0}, \quad \boldsymbol{\mu_2} = \mathbf{X}\boldsymbol{\beta}, \quad \mathbf{V}_{\mathbf{x_1}\mathbf{x_2}} = \mathbf{G}\mathbf{Z}^T, \quad \mathbf{V}_{\mathbf{x_2}\mathbf{x_2}} = \mathbf{V}$$

yielding the BLUP of $\mathbf{u}$ as

$$\begin{aligned}\widehat{\mathbf{u}} &= \boldsymbol{\mu_1} + \mathbf{V}_{\mathbf{x_1}\mathbf{x_2}}\mathbf{V}_{\mathbf{x_2}\mathbf{x_2}}^{-1}(\mathbf{x_2} - \boldsymbol{\mu_2})\\ &= \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)\end{aligned} \tag{9.22c}$$

(Henderson 1963). Further, applying Equation 8.28 gives the variance of $\widehat{\mathbf{u}}$ as

$$\mathbf{G} - \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \tag{9.22d}$$

Suppose that we can express $\mathbf{G}$ as $\sigma_G^2\,\mathbf{A}$ and $\mathbf{R}$ as $\sigma_E^2\,\mathbf{B}$. While one might suspect that two variance components (σ_G^2 and σ_E^2) are required to apply Equation 9.22c. In actuality, we only require a single parameter, the ratio σ_E^2/σ_G^2. To see this, first note that

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} = \sigma_G^2[\,\mathbf{Z}\mathbf{A}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{B}\,] \tag{9.23a}$$

Hence,

$$\begin{aligned}\mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1} &= (\sigma_G^2\mathbf{A})\,\mathbf{Z}^T\sigma_G^{-2}[\,\mathbf{Z}\mathbf{A}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{B}\,]^{-1}\\ &= \mathbf{A}\,\mathbf{Z}^T[\,\mathbf{Z}\mathbf{A}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{B}\,]^{-1}\end{aligned} \tag{9.23b}$$

In a similar fashion,

$$\begin{aligned}\boldsymbol{\beta} &= \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}\\ &= \left(\mathbf{X}^T\sigma_G^{-2}[\,\mathbf{Z}\mathbf{A}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{B}\,]^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\sigma_G^{-2}[\,\mathbf{Z}\mathbf{A}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{B}\,]^{-1}\mathbf{y}\\ &= \left(\mathbf{X}^T[\,\mathbf{Z}\mathbf{A}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{B}\,]^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T[\,\mathbf{Z}\mathbf{A}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{B}\,]^{-1}\mathbf{y}\end{aligned} \tag{9.23c}$$

showing that the BLUE also only depends on the variance ratio (σ_E^2/σ_G^2).

Finally, a few comments on an important feature, **shrinkage**, of BLUPs are in order. Equation 9.22c shows why we speak of *predicting* random effects, as their values are *predicted from a regression*. Consider the ith observation, which we can express as

$$y_i - \widehat{y}_i = u_i + e_i \tag{9.24a}$$

The central question becomes of much of the deviation between the observed y_i and predicted $\widehat{y}_i$ values is due to the random effect u_i versus the residual e_i. Intuition suggests that if most of the variance is in u, then most of the deviation is due to u_i, while if most of the variation is from residual effects, then very little of the deviation is from u_i. We can more formally illustrate this by supposing only a single observation, so that $\mathbf{Z} = (1), \mathbf{G} = (\sigma_G^2)$, $\mathbf{R} = (\sigma_E^2)$, and $\mathbf{V} = (\sigma_G^2 + \sigma_E^2)$. In this setting, Equation 9.22c reduces to

$$\widehat{u} = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}\,(y - \widehat{y}\,) = \begin{cases} \simeq 0 & \text{for } \sigma_E^2 \gg \sigma_G^2 \\ \simeq y - \widehat{y} & \text{for } \sigma_G^2 \gg \sigma_E^2 \end{cases} \tag{9.24b}$$

The ratio of σ_G^2 to the total variance $(\sigma_G^2 + \sigma_E^2)$ is the **shrinkage factor** for this random effect. If this ratio is close to one, there is very little shrinkage, with $u_i = y_i - \widehat{y}_i$, which is essentially the estimate obtained by treating u_i as a fixed effect. Conversely, if this ratio is near zero, them almost all of the deviation is due to the residual error, with u_i being assigned a value close to its mean (which, by construction, is zero). Example 9.13 further expands on these ideas.

Example 9.11. What are the BLUP values for the sire effects (u_1, u_2, u_3) in Example 9.10? In order to proceed, we require the covariance matrices for the sire effects and the residual errors. We assume that the residual variances within both environments are the same (σ_E^2), so $\mathbf{R} = \sigma_E^2\,\mathbf{I}$, where $\mathbf{I}$ is the 6×6 identity matrix. Assuming that all three sires are unrelated and drawn from the same population, $\mathbf{G} = \sigma_S^2\,\mathbf{I}$, where $\mathbf{I}$ is the 3×3 identity matrix and σ_S^2 is the variance of sire effects. Assuming only additive genetic variance, the sire effects (breeding values) are half the sires' additive genetic values (Chapters 7 and 21). Thus, because the sires are sampled randomly from an outbred base population, $\sigma_S^2 = \sigma_A^2/4$, where σ_A^2 is the additive genetic variance. Assume that $\sigma_A^2 = 8$ and $\sigma_E^2 = 6$ for a heritability (assuming only additive variances) of $h^2 = \sigma_A^2/(\sigma_A^2 + \sigma_e^2) = 8/(6 + 8) = 0.57$. With these values, the covariance matrix $\mathbf{V}$ for the vector of observations $\mathbf{y}$ is given by $\mathbf{ZGZ}^T + \mathbf{R}$, or

$$\mathbf{V} = \frac{8}{4}\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} + 6\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 8 & 2 & 0 & 0 & 0 & 0 \\ 2 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 2 & 0 & 0 \\ 0 & 0 & 2 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 2 \\ 0 & 0 & 0 & 0 & 2 & 8 \end{pmatrix}, \quad \text{giving} \quad \mathbf{V}^{-1} = \frac{1}{30}\cdot\begin{pmatrix} 4 & -1 & 0 & 0 & 0 & 0 \\ -1 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & -1 & 0 & 0 \\ 0 & 0 & -1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -1 \\ 0 & 0 & 0 & 0 & -1 & 4 \end{pmatrix}$$

Using this result, a few simple matrix calculations give

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} = \frac{1}{18}\begin{pmatrix} 148 \\ 235 \end{pmatrix}$$

and

$$\widehat{\mathbf{u}} = \begin{pmatrix} \widehat{u}_1 \\ \widehat{u}_2 \\ \widehat{u}_3 \end{pmatrix} = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) = \frac{1}{18}\begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

Example 9.12. As a prelude to associate analysis (Chapter 19), the effects on a trait from different genotypes at a single candidate gene are often estimated by ordinary least squares (OLS), using the model

$$y_{ij} = g_i + e_{ij} \tag{9.25a}$$

where y_{ij} is the observed phenotype of the jth individual of genotype i, g_i is the mean genotypic value for the ith genotype at the locus of interest, and e_{ij} is a residual deviation assumed to be independently distributed among individuals. While this model may be reasonable for a random collection of individuals from a large population, when some sampled individuals are relatives, the sharing of alleles at other loci influencing the trait will induce correlations between residuals. If this is the case, OLS analysis can produce biased estimates of the focal QTL effects. When one of the QTL genotypes is very rare, as is often the case, the sampled individuals may be intentionally selected from the same pedigree, so the problem of bias is not trivial.

Use of a mixed model provides a means of accounting for associations among background QTLs in realtives in a way that eliminates bias in estimates of QTL effects. If the relatives in question share only additive effects (as in a pedigree with no full sibs or double first cousins, or when there is no nonadditive gene action), the correlations among residuals are accounted for by the additive genetic relationship matrix $\mathbf{A}$, where A_{ij} is twice the coefficient of coancestry, $2\Theta_{ij}$ (Equation 7.11a). When full sibs are included and dominance is present at background QTLs, both $\mathbf{A}$ and a dominance relationship matrix (Chapter 29) are required. Here we assume that all of the background genetic effects are additive, in which case the simplest mixed model can be applied,

$$y_{ij} = g_i + a_{ij} + e_{ij} \tag{9.25b}$$

with the contribution from the different single-locus genotypes (g_i) being treated as fixed effects. The additive genetic background effects (a_{ij}) and the residual environmental deviations (e_{ij}) are treated as random effects, both with expected values equal to zero, and with respective variances σ_A^2 and σ_E^2. Note that σ_A^2 is the background additive genetic variance for the trait in excess of that caused by the focal QTL. In matrix form,

$$\mathbf{y} = \mathbf{X}\mathbf{g} + \mathbf{Z}\mathbf{a} + \mathbf{e} \tag{9.25c}$$

If there is a single observation for each individual, as we assume below, then $\mathbf{Z} = \mathbf{I}$ and the covariance matrix for the vector of observations $(\mathbf{y})$ becomes

$$\mathbf{V} = \sigma_A^2\,\mathbf{A} + \sigma_E^2\,\mathbf{I} \tag{9.25d}$$

Thus, the covariance between the residual errors of two individuals (i and j) is just $2\Theta_{ij}\sigma_A^2$, while the variance of individual errors is $\sigma_A^2 + \sigma_E^2$. The error in using OLS to estimate single gene effects is that $\mathbf{A}$ is (incorrectly) assumed to equal an identity matrix, so that $\mathbf{V}$ is incorrectly assumed to be diagonal.

From Equation 9.21, the estimates of the QTL means are given by

$$\widehat{\mathbf{g}} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \tag{9.25e}$$

Kennedy et al. (1992) showed that mixed-model estimates of QTL effects are much more reliable than OLS estimates, especially in small selected populations. Building on this approach, several authors (Hoeschele 1988; Hofer and Kennedy 1993; Kinghorn et al. 1993) have proposed BLUP-based segregation analysis (Chapter 15) for estimating the effects of an unknown major gene. Here the elements in the design matrix $\mathbf{X}$ associated with g_i are probabilistic estimates for the major-locus genotypes of each individual. As we develop in Chapter

19, the principles of this example, along with those of Example 8.13, give the structure of mixed models used in association analysis.

Example 9.13. Equations 9.21 and 9.22c can be used to provide significant insight into the implications of treating a factor as fixed versus random. Suppose we measure two replicate plots for each of three inbred lines in a single randomized block, so that the simplest model is

$$y_{ij} = L_i + e_{ij} \tag{9.26a}$$

where y_{ij} is the value of the jth plot of line i. Suppose that the plot values were 6 and 8 for line 1, 10 and 12 for line 2, and 14 and 16 for line 3. Treating the line values as fixed, the resulting GLM, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, has

$$\mathbf{y} = \begin{pmatrix} 6 \\ 8 \\ 10 \\ 12 \\ 14 \\ 16 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} \tag{9.26b}$$

Assuming OLS residuals, Equation 9.9a gives the OLS BLUEs for the line values as

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \begin{pmatrix} 7 \\ 11 \\ 15 \end{pmatrix}$$

Now use this exact same data set, but treat the line values as random. To provide as fair a comparison as possible, assume the lines are unrelated (so that information is not borrowed from correlated observations). The covariance structure for the vector of random line effects is thus $\sigma_G^2\,\mathbf{I}$. The model now becomes

$$y_{ij} = \mu + \ell_i + e_{ij} \tag{9.26c}$$

where μ is the overall mean, and ℓ_i the realization for line i (giving the line value as $L_i = \mu + \ell_i$). Again, assume OLS residuals. The resulting components of the mixed model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, become

$$\mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{\beta} = (\,\mu\,), \qquad \mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{u} = \begin{pmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{pmatrix} \tag{9.26d}$$

with a covariance structure of

$$\text{Var}\begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \sigma_G^2\,\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_E^2\,\mathbf{I} \end{pmatrix}$$

Using Equation 9.23a, these assumptions yield

$$\begin{aligned} \mathbf{V} &= \sigma_G^2[\,\mathbf{Z}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{I}\,] \\ &= \sigma_G^2\left[\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} + \sigma_E^2/\sigma_G^2 \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}\right] \end{aligned} \tag{9.26e}$$

One can show for this $\mathbf{V}$ that all values of σ_E^2/σ_G^2 yield

$$\widehat{\boldsymbol{\beta}} = \widehat{\mu} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} = (11)$$

Equations 9.22c and 9.23b gives the solution for the BLUPs of line values as

$$\widehat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) = \mathbf{Z}^T[\,\mathbf{Z}\mathbf{Z}^T + (\sigma_E^2/\sigma_G^2)\,\mathbf{I}\,]^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)$$

Consider σ_E^2/σ_G^2 values of 500, 5, 1, 1/5, and $\simeq 0$. Noting that $H^2 = 1/(1 + \sigma_E^2/\sigma_G^2)$, these correspond to broad-sense heritability values of 0.002, 0.17, 0.5, 0.82, and $\simeq$1, respectively, with corresponding BLUPs of

$$\widehat{\mathbf{u}} = \begin{pmatrix} -0.016 \\ 0 \\ 0.016 \end{pmatrix}, \quad \begin{pmatrix} -1.14 \\ 0 \\ 1.14 \end{pmatrix}, \quad \begin{pmatrix} -2.67 \\ 0 \\ 2.67 \end{pmatrix}, \quad \begin{pmatrix} -3.64 \\ 0 \\ 3.64 \end{pmatrix}, \quad \begin{pmatrix} -4.00 \\ 0 \\ 4.00 \end{pmatrix}$$

These solutions highlight the property of shrinkage that arises with the prediction of random effects. When H^2 is near zero, so that the phenotype of a line is a very poor predictor of its genetic value, the BLUPs are regressed back to near the mean (which is zero for a random effect, for an L value of $\widehat{\mu} = 11$). Conversely, when H^2 is near one, we essentially recover the fixed-effect solutions.

Hence, there is a tradeoff with using random effects. On one hand, they shrink observed deviations back towards the mean, which is not done by BLUEs. On the other hand, they burrow information from correlated observations that are not used by BLUEs.

Henderson's Mixed Model Equations and MM Standard Errors

Computation of Equations 9.21 and 9.22c requires the inverse of the covariance matrix $\mathbf{V}$. In the above toy examples, $\mathbf{V}^{-1}$ was not difficult to obtain. However, when $\mathbf{y}$ contains many hundreds-of-thousands of observations, as is commonly the case in cattle breeding, the computation of $\mathbf{V}^{-1}$ can be challenging. As a way around this problem, Henderson (1950, 1963, 1973, 1984a) offered a more compact method for jointly obtaining $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ in the form of his **mixed-model equations** (MME),

$$\begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \qquad (9.27)$$

While these expressions may look considerably more complicated than Equations 9.21 and 9.22c, $\mathbf{R}^{-1}$ and $\mathbf{G}^{-1}$ are trivial to obtain if $\mathbf{R}$ and $\mathbf{G}$ are diagonal, and hence the submatrices in Equation 9.27 are much easier to compute than $\mathbf{V}^{-1}$. A second advantage of Equation 9.27 can be seen by considering the dimensionality of the matrix on the left hand side. Recalling that $\mathbf{X}$ and $\mathbf{Z}$ are $n \times p$ and $n \times q$ respectively, $\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}$ is $p \times p$, $\mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z}$ is $p \times q$, and $\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}$ is $q \times q$. Thus, the matrix that needs to be inverted to obtain the solution for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ is of order $(p+q) \times (p+q)$, which is often considerably less than the dimensionality of $\mathbf{V}$ (an $n \times n$ matrix). Further, one can solve Equation 9.27 by Gaussian elimination, bypassing matrix inversion.

Although there are several ways to derive the mixed-model equations (Robinson 1991; a derivation can be found in WL Example A6.5), Henderson (1950) originally obtained them by assuming that the covariance matrices $\mathbf{G}$ and $\mathbf{R}$ are known and that the densities of the vectors $\mathbf{u}$ and $\mathbf{e}$ are each multivariate normal with no correlations between them. Equation 9.27 then yields the maximum likelihood estimates of the fixed and random effects. Henderson (1963) later showed that the mixed-model equations do not actually depend on

normality, and that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ are BLUE and BLUP, respectively, under general conditions provided the variances are known.

Example 9.14. Using the values from Examples 9.10 and 9.11, we find that

$$\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} = \frac{1}{6}\begin{pmatrix}4 & 0\\ 0 & 2\end{pmatrix}, \qquad \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} = \left(\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X}\right)^T = \frac{1}{6}\begin{pmatrix}1 & 2 & 1\\ 1 & 0 & 1\end{pmatrix}$$

$$\mathbf{G}^{-1} + \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} = \frac{5}{6}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}, \quad \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} = \frac{1}{6}\begin{pmatrix}33\\ 26\end{pmatrix}, \quad \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} = \frac{1}{6}\begin{pmatrix}21\\ 17\\ 21\end{pmatrix}$$

Thus, after factoring out $1/6$ from both sides, the mixed-model equations for these data become

$$\begin{pmatrix}4 & 0 & 1 & 2 & 1\\ 0 & 2 & 1 & 0 & 1\\ 1 & 1 & 5 & 0 & 0\\ 2 & 0 & 0 & 5 & 0\\ 1 & 1 & 0 & 0 & 5\end{pmatrix}\begin{pmatrix}\widehat{\beta}_1\\ \widehat{\beta}_2\\ \widehat{u}_1\\ \widehat{u}_2\\ \widehat{u}_3\end{pmatrix} = \begin{pmatrix}33\\ 26\\ 21\\ 17\\ 21\end{pmatrix}$$

Taking the inverse gives the solution

$$\begin{pmatrix}\widehat{\beta}_1\\ \widehat{\beta}_2\\ \widehat{u}_1\\ \widehat{u}_2\\ \widehat{u}_3\end{pmatrix} = \frac{1}{270}\begin{pmatrix}100 & 25 & -25 & -40 & -25\\ 25 & 175 & -40 & -10 & -40\\ -25 & -40 & 67 & 10 & 13\\ -40 & -10 & 10 & 70 & 10\\ -25 & -40 & 13 & 10 & 67\end{pmatrix}\begin{pmatrix}33\\ 26\\ 21\\ 17\\ 21\end{pmatrix} = \frac{1}{18}\begin{pmatrix}148\\ 235\\ -1\\ 2\\ -1\end{pmatrix}$$

which is identical to the results obtained in Example 9.11.

A relatively straightforward extension of Henderson's mixed-model equations provides estimates of the standard errors of the fixed and random effects. Write the inverse of the **left hand side** (**LHS**) matrix in Equation 9.27 as

$$\begin{pmatrix}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\end{pmatrix}^{-1} = \begin{pmatrix}\mathbf{C}_{11} & \mathbf{C}_{12}\\ \mathbf{C}_{12}^T & \mathbf{C}_{22}\end{pmatrix} \tag{9.28}$$

where $\mathbf{C}_{11}$, $\mathbf{C}_{12}$, and $\mathbf{C}_{22}$ are, respectively, $p \times p$, $p \times q$, and $q \times q$ submatrices. Using this notation, Henderson (1975) showed that the sampling covariance matrix for the BLUE of $\boldsymbol{\beta}$ is

$$\boldsymbol{\sigma}(\widehat{\boldsymbol{\beta}}) = \mathbf{C}_{11} \tag{9.29a}$$

while the sampling covariance matrix of the prediction errors $(\widehat{\mathbf{u}} - \mathbf{u})$ is

$$\boldsymbol{\sigma}(\,\widehat{\mathbf{u}} - \mathbf{u}\,) = \mathbf{C}_{22} \tag{9.29b}$$

and, finally, the sampling covariance of estimated effects and prediction errors is

$$\boldsymbol{\sigma}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{u}} - \mathbf{u}\,) = \mathbf{C}_{12} \tag{9.29c}$$

(We consider $\widehat{\mathbf{u}} - \mathbf{u}$ rather than $\widehat{\mathbf{u}}$ as the latter includes variance from both the prediction error and the random effects $\mathbf{u}$ themselves.) The standard errors of the fixed and random

effects are obtained, respectively, as the square roots of the diagonal elements of $\mathbf{C}_{11}$ and $\mathbf{C}_{22}$. For very large animal breeding designs where the inverse of the MME matrix may be difficult to compute, Meyer (1989a) presents methods for approximating the diagonal elements of the inverse of this matrix (and hence the standard errors).

Example 9.15. Consider the mixed-model equation from Example 9.14. Here for the fixed factors β_1, β_2 and the random effects u_1, u_2, u_3, the inverse of the LHS coefficient matrix is

$$\begin{pmatrix} 4 & 0 & \vdots & 1 & 2 & 1 \\ 0 & 2 & & 1 & 0 & 1 \\ \cdots & & \cdots & & \cdots & \\ 1 & 1 & & 5 & 0 & 0 \\ 2 & 0 & \vdots & 0 & 5 & 0 \\ 1 & 1 & & 0 & 0 & 5 \end{pmatrix}^{-1} = \frac{1}{270} \begin{pmatrix} 100 & 25 & \vdots & -25 & -40 & -25 \\ 25 & 175 & & -40 & -10 & -40 \\ \cdots & & \cdots & & \cdots & \\ -25 & -40 & & 67 & 10 & 13 \\ -40 & -10 & \vdots & 10 & 70 & 10 \\ -25 & -40 & & 13 & 10 & 67 \end{pmatrix}$$

Hence,

$$\mathbf{C}_{11} = \frac{1}{270} \begin{pmatrix} 100 & 25 \\ 25 & 175 \end{pmatrix} \quad \text{and} \quad \mathbf{C}_{22} = \frac{1}{270} \begin{pmatrix} 67 & 10 & 13 \\ 10 & 70 & 10 \\ 13 & 10 & 67 \end{pmatrix}$$

so that, for example,

$$\sigma^2(\widehat{\beta}_1) = \frac{100}{270}, \quad \sigma^2(\widehat{\beta}_2) = \frac{175}{270}, \quad \sigma(\widehat{\beta}_1, \widehat{\beta}_2) = \frac{25}{270}$$

and, likewise,

$$\sigma^2(\widehat{u}_2 - u_2) = \frac{70}{270}, \quad \sigma(\widehat{u}_1 - u_1, \widehat{u}_3 - u_3) = \frac{13}{270}, \quad \text{and so on.}$$

REML Estimation of Variance Components

Finally, as noted above, the practical application of Equations 9.21, 9.22c, and 9.27 requires that the variance components be known. Thus, prior to a BLUP analysis, these need to be estimated by ANOVA (Chapter 21) or **restricted maximum likelihood** (**REML**). REML is closely related to BLUP, with (roughly speaking) REML estimates obtained from iterating and updating BLUP estimates until there is suitable convergence (Chapter 30). REML maximizes that part of the likelihood function that is unaffected by fixed effects (Patterson and Thompson 1971). Harville (1977) coined the term **restricted ML**, but Thompson (2008) noted that REML maximizes a *residual* likelihood, and hence preferred the term **residual maximum likelihood**. One advantage of REML estimates (over those obtained by other estimation procedures) is that they are unbiased by the estimates of fixed effects (Patterson and Thompson 1971). Chapter 30 examines REML in some detail.

Thus, BLUPs are usually obtained by a two-stage approach: variance components are first obtained by REML and then these estimates are used to obtain BLUPs. This two-stage approach is called **empirical BLUP** or **REML/BLUP** (Sorensen and Kennedy 1986; Kennedy and Sorensen 1988; Harville 1990). Kackar and Harville (1981) and Gianola et al. (1986, 1988) showed that using REML estimation does not result in biased values for BLUPs, but that the resulting predictors may not be "best" (there may be other linear predictors with smaller mean-squared errors). While Chapter 30 examines large-sample approximations for the uncertainty in REML variance estimates, these cannot be easily translated into how much additional uncertainty is introduced into BLUPs by using REML estimates in place of their true values. A more formal procedure is to use a Bayesian approach (Appendices

7 and 8), wherein one draws a value of the variances from their posterior distribution, and then computes the BLUPs given this value. Repeating this procedure generates a posterior distribution for BLUPs that fully incorporates any errors introduced by using estimated, instead of true, variances.

Literature Cited

Aitken, A. C. 1935. On least squares and linear combination of observations. *Proc. Royal Soc. Edinburgh A* 55: 42–47. [9]

Eisenhart, C. 1947. The assumptions underlying the analysis of variance. *Biometrics* 3: 1–21. [9]

Gianola, D., J. L. Foulley, and R. L. Fernando. 1986. Prediction of breeding values when variances are not know. *Genet. Sel. Evol.* 18: 485–498. [9]

Gianola, D., S. Im, and R. L. Fernando. 1988. Prediction of breeding values under Henderson's selection model: a revisitation. *J. Dairy Sci.* 71: 2790–2798. [9]

Goldberger, A. S. 1962. Best linear unbiased predictors in the generalized linear regression model. *J. Am. Stat. Assoc.* 57: 369–375. [9]

Harville, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* 72: 320–338. [9]

Harville, D. A. 1990. BLUP (Best Linear Unbiased Prediction) and beyond. *In* D. Gianola and K. Hammond (eds.), *Advances in statistical methods for genetic improvement of livestock*, pp. 239–276. Springer-Verlag, New York, NY. [9]

Henderson, C. R. 1949. Estimates of changes in herd environment. *J. Dairy Sci.* 32: 706. [9]

Henderson, C. R. 1950. Estimation of genetic parameters. *Ann. Math. Stat.* 21: 309–310. [9]

Henderson, C. R. 1963. Selection index and the expected genetic advance *In* W. D. Hanson and H. F. Robinson (eds.), *Statistical genetics and plant breeding*, pp. 141–163. Natl. Acad. Sci., Natl. Res. Council Publ. No. 982, Washington, D. C. [9].

Henderson, C. R. 1973. Sire evaluation and genetic trends. *In* Proceedings of the animal breeding and genetics symposium in honor of Dr. J. L. Lush, pp. 10–41. American Society of Animal Science, Champaign, IL. [9]

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31: 423–447. [9]

Henderson, C. R. 1984a. *Applications of linear models in animal breeding*. Univ. Guelph, Guelph, Ontario. [9]

Hoaglin, D. C., and R. E. Welsch. 1978. The Hat Matrix in regression and ANOVA. *Am. Stat.* 32: 17–22. [9]

Hoeschele, I. 1988. Statistical techniques for detection of major genes in animal breeding data. *Theor. Appl. Genet.* 76: 311–319. [9]

Hofer, A., and B. W. Kennedy. 1993. Genetic evaluation for a quantitative trait controlled by polygenes and a major locus with genotypes not or only partly known. *Genet. Sel. Evol.* 25: 537–555. [9]

Kackar, R. N., and D. A. Harville. 1981. Unbiasedness of two-stage estimation and prediction for mixed linear models. *Comm. Stat. Theor. Meth.* A10: 1249–1261. [9]

Kennedy, B. W., M. Quinton, and J. A. M. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* 70: 2000–2012. [9]

Kennedy, B. W., and D. A. Sorensen. 1988. Properties of mixed-model methods for prediction of genetic merit. *In* B. S. Weir, E. J. Eisen, M. M. Goodman, and G. Namkoong (eds.), *Proceedings of the second international conference on quantitative genetics,* pp. 91–103. Sinauer Assoc., Sunderland, MA. [9]

Kinghorn, B. P., B. W. Kennedy, and C. Smith. 1993. A method for screening for genes of major effect. *Genetics* 134: 351–360. [9]

Meyer, K. 1989a. Approximate accuracy of genetic evaluation under an animal model. *Livestock Prod. Sci.* 21: 87–100. [9]

Patterson, H. D., and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 58: 545–554. [9]

Robinson, G. K. 1991. That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* 6: 15–51. [9]

Sorensen, D. A., and B. W. Kennedy. 1986. Analysis of selection experiments using mixed model methodology. *J. Anim. Sci.* 63: 245–258. [19]

Sorensen , D. A., and B. W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. *J. Anim. Sci.* 59: 1213–1233. [9]

Thompson, R. 2008. Estimation of quantitative genetic parameters. *Proc. Roy. Soc. Lond. B* 275: 679–686. [9]