Chapter 12
Mixed Models and Variance Components

Traditionally, linear models have been divided into three categories: *fixed effects models*, *random effects models*, and *mixed models*. The categorization depends on whether the β vector in $Y = X\beta + e$ is fixed, random, or has both fixed and random elements. Random effects models always assume that there is a fixed overall mean for observations, so random effects models are actually mixed.

Variance components are the variances of the random elements of β. Sections 1 through 3 discuss mixed models in general and prediction for mixed models. Sections 4 through 9 present methods of estimation for variance components. Section 10 examines exact tests for variance components. Section 11 uses the ideas of the chapter to develop the interblock analysis for balanced incomplete block designs. Searle, Casella, and McCulloch (1992) give an extensive discussion of variance component estimation. Khuri, Mathew, and Sinha (1998) give an extensive discussion of testing in mixed models.

The methods considered in this chapter are presented in terms of fitting general linear models. In many special cases, considerable simplification results. For example, the RCB models of (11.1.5) and Exercise 11.4, the split plot model (11.3.1), and the subsampling model (11.4.2) are all mixed models with very special structures.

12.1 Mixed Models

The *mixed model* is a linear model in which some of the parameters, instead of being fixed effects, are random. The model can be written

$$Y = X\beta + Z\gamma + e, \tag{1}$$

where X and Z are known matrices, β is an unobservable vector of fixed effects, and γ is an unobservable vector of random effects with $\mathrm{E}(\gamma) = 0$, $\mathrm{Cov}(\gamma) = D$, and $\mathrm{Cov}(\gamma, e) = 0$. Let $\mathrm{Cov}(e) = R$.

R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer Texts in Statistics, DOI 10.1007/978-1-4419-9816-3_12,

For estimation of the fixed effects β, the mixed model can be written as a general Gauss–Markov model. Write

$$V = \mathrm{Cov}(Y) = \mathrm{Cov}(Z\gamma + e) = ZDZ' + R.$$

V is assumed to be nonsingular. Model (1) is equivalent to

$$Y = X\beta + \xi, \quad \mathrm{E}(\xi) = 0, \quad \mathrm{Cov}(\xi) = V.$$

The BLUE of $X\beta$ can be found using the theory for general Gauss–Markov models. Unfortunately, finding the BLUE requires knowledge (at least up to a constant multiple) of V. This is rarely available. Currently, the best procedure available for estimating $X\beta$ is to estimate V and then act as if the estimate is the real value of V. In other words, if V is estimated with $\hat{V}$, then $X\beta$ is estimated with

$$X\hat{\beta} = X\left[X'\hat{V}^{-1}X\right]^{-}X'\hat{V}^{-1}Y. \tag{2}$$

If $\hat{V}$ is close to V, then the estimate of $X\beta$ should be close to the BLUE of $X\beta$. Corresponding standard errors will tend to be underestimated, cf. Eaton (1985), Harville (1985), or Christensen (2001, Section 6.5).

Estimation of V is obviously crucial for the estimation of the fixed effects. It is also frequently of interest in its own right. Estimation of V is indirectly referred to as *variance component estimation*. We discuss several approaches to the problem of estimating V.

If γ were a fixed effect, we would be interested in estimating estimable functions like $\lambda'\gamma$. Since γ is random, we cannot estimate $\lambda'\gamma$, but we can predict $\lambda'\gamma$ from the observable vector Y. The theory of Subsection 6.3.4 provides a best linear predictor, but requires knowledge of $\mathrm{E}(\lambda'\gamma)$, $\mathrm{Cov}(\lambda'\gamma, Y)$, $\mathrm{E}(Y)$, and $\mathrm{Cov}(Y)$. We know that $\mathrm{E}(\lambda'\gamma) = 0$. As will be seen in Section 2, if we assume that D and R are known, our only problem in predicting $\lambda'\gamma$ is not knowing $\mathrm{E}(Y) = X\beta$. We will not be able to find a best linear predictor, but we will find the *best linear unbiased predictor (BLUP)* of $\lambda'\gamma$. Of course, D and R are not known, but we will have a theory for prediction of $\lambda'\gamma$ that is the equivalent of BLUE estimation. With good estimates of D and R provided by variance component techniques, we can get predictions of $\lambda'\gamma$ that are close to the BLUP.

Sometimes we would like to obtain an estimate of $X\beta$ without going to the trouble and expense of finding V^{-1}, which is needed to apply equation (2). One simple estimate is the least squares estimate, MY. This gives an unbiased estimate of $X\beta$, but ignores the existence of the random effects. An alternative method is to fit the model

$$Y = X\beta + Z\delta + Z\gamma + e, \quad R = \sigma^2 I, \tag{3}$$

where δ is a vector of fixed effects corresponding to the vector of random effects γ. In model (3), there is no hope of estimating variance components because the fixed effects δ and the random effects γ are completely confounded. However, it is easily seen that $C(V[X,Z]) \subset C(X,Z)$, so by Theorem 10.4.5, least squares estimates are

BLUEs in model (3). To see that $C(V[X,Z]) \subset C(X,Z)$, observe that

$$\begin{aligned} C(V[X,Z]) &= C(\{\sigma^2 I + ZDZ'\}[X,Z]) \\ &= C(\sigma^2 X + ZDZ'X, \sigma^2 Z + ZDZ'Z) \\ &\subset C(X,Z). \end{aligned}$$

From Chapter 9, a least squares estimate of β is

$$\hat{\beta} = \left[X'(I - M_Z)X\right]^{-} X'(I - M_Z)Y, \tag{4}$$

where $M_Z = Z(Z'Z)^{-}Z'$. Although estimates obtained using (4) are not typically BLUEs for model (1), since model (3) is a larger model than model (1), the estimates should be reasonable. The only serious problem with using equation (4) is that it is not clear which functions of β are estimable in model (3).

12.2 Best Linear Unbiased Prediction

In this section we consider the general theory of best linear unbiased prediction. In a series of examples, this theory will be used to examine prediction in standard linear model theory, Kriging of spatial data, and prediction of random effects in mixed models. The placement of this section in Chapter 12 is obviously based on the last of these applications.

Consider a set of random variables y_i, $i = 0, 1, \ldots, n$. We want to use $y_1, \ldots, y_n$ to predict y_0. Let $Y = (y_1, \ldots, y_n)'$. In Section 6.3 it was shown that the best predictor (BP) of y_0 is $\mathrm{E}(y_0|Y)$. Typically, the joint distribution of y_0 and Y is not available, so $\mathrm{E}(y_0|Y)$ cannot be found. However, if the means and covariances of the y_is are available, then we can find the best linear predictor (BLP) of y_0. Let $\mathrm{Cov}(Y) = V$, $\mathrm{Cov}(Y, y_0) = V_{y0}$, $\mathrm{E}(y_i) = \mu_i$, $i = 0, 1, \ldots, n$, and $\mu = (\mu_1, \ldots, \mu_n)'$. Again from Subsection 6.3.4, the BLP of y_0 is

$$\hat{E}(y_0|Y) \equiv \mu_0 + \delta_*'(Y - \mu), \tag{1}$$

where δ_* satisfies $V\delta_* = V_{y0}$.

We now want to weaken the assumption that μ and μ_0 are known. Since the prediction is based on Y and there are as many unknown parameters in μ as there are observations in Y, we need to impose some structure on the mean vector μ before we can generalize the theory. This will be done by specifying a linear model for Y. Since y_0 is being predicted and has not been observed, it is necessary either to know μ_0 or to know that μ_0 is related to μ in a specified manner. In the theory below, it will be assumed that μ_0 is related to the linear model for Y.

Suppose that a vector of known concomitant variables $x_i' = (x_{i1}, \ldots, x_{ip})$ is associated with each random observation y_i, $i = 0, \ldots, n$. We impose structure on the

μ_is by assuming that $\mu_i = x_i'\beta$ for some vector of unknown parameters β and all $i = 0, 1, \ldots, n$.

We can now reset our notation in terms of linear model theory. Let

$$X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}.$$

The observed vector Y satisfies the linear model

$$Y = X\beta + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = V. \tag{2}$$

With this additional structure, the BLP of y_0 given in (1) becomes

$$\hat{E}(y_0|Y) = x_0'\beta + \delta_*'(Y - X\beta), \tag{3}$$

where again $V\delta_* = V_{y0}$.

The standard assumption that μ and μ_0 are known now amounts to the assumption that β is known. It is this assumption that we renounce. By weakening the assumptions, we consequently weaken the predictor. If β is known, we can find the best linear predictor. When β is unknown, we find the best linear unbiased predictor.

Before proceeding, a technical detail must be mentioned. To satisfy estimability conditions, we need to assume that $x_0' = \rho'X$ for some vector ρ. Frequently, model (2) will be a regression model, so choosing $\rho' = x_0'(X'X)^{-1}X'$ will suffice. In applications to mixed models, $x_0' = 0$, so $\rho = 0$ will suffice.

Definition 12.2.1. A *predictor* $f(Y)$ of y_0 is said to be *unbiased* if

$$\mathrm{E}[f(Y)] = \mathrm{E}(y_0).$$

Definition 12.2.2. $a_0 + a'Y$ is a *best linear unbiased predictor* of y_0 if $a_0 + a'Y$ is unbiased and if, for any other unbiased predictor $b_0 + b'Y$,

$$\mathrm{E}\left[y_0 - a_0 - a'Y\right]^2 \leq \mathrm{E}\left[y_0 - b_0 - b'Y\right]^2.$$

Theorem 12.2.3. The best linear unbiased predictor of y_0 is $x_0'\hat{\beta} + \delta_*'(Y - X\hat{\beta})$, where $V\delta_* = V_{y0}$ and $X\hat{\beta}$ is a BLUE of $X\beta$.

PROOF. The technique of the proof is to change the prediction problem into an estimation problem and then to use the theory of best linear unbiased estimation. Consider an arbitrary linear unbiased predictor of y_0, say $b_0 + b'Y$. By Proposition 6.3.3,

$$\mathrm{E}[y_0 - b_0 - b'Y]^2 = \mathrm{E}[y_0 - \hat{E}(y_0|Y)]^2 + \mathrm{E}[\hat{E}(y_0|Y) - b_0 - b'Y]^2;$$

so it is enough to find $b_0 + b'Y$ that minimizes $\mathrm{E}[\hat{E}(y_0|Y) - b_0 - b'Y]^2$.

From the definition of $\hat{E}(y_0|Y)$ and unbiasedness of $b_0 + b'Y$, we have

$$0 = \mathrm{E}[\hat{E}(y_0|Y) - b_0 - b'Y].$$

Substituting from equation (3) gives

$$0 = \mathrm{E}[x_0'\beta + \delta_*'(Y - X\beta) - b_0 - b'Y].$$

This relationship holds if and only if $b_0 + (b - \delta_*)'Y$ is a linear unbiased estimate of $x_0'\beta - \delta_*'X\beta$. By Proposition 2.1.9,

$$b_0 = 0;$$

so the term we are trying to minimize is

$$\begin{aligned}\mathrm{E}[\hat{E}(y_0|Y) - b_0 - b'Y]^2 &= \mathrm{E}[(b - \delta_*)'Y - (x_0'\beta - \delta_*'X\beta)]^2 \\ &= \mathrm{Var}\left[(b - \delta_*)'Y\right].\end{aligned}$$

Because $(b - \delta_*)'Y$ is a linear unbiased estimate, to minimize the variance choose b so that $(b - \delta_*)'Y = x_0'\hat{\beta} - \delta_*'X\hat{\beta}$ is a BLUE of $x_0'\beta - \delta_*'X\beta$. It follows that the best linear unbiased predictor of y_0 is

$$b'Y = x_0'\hat{\beta} + \delta_*'(Y - X\hat{\beta}). \qquad \square$$

It should not be overlooked that both $\hat{\beta}$ and δ_* depend crucially on V. When all inverses exist, $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$ and $\delta_* = V^{-1}V_{y0}$. However, this proof remains valid when either inverse does not exist.

It is also of value to know the prediction variance of the BLUP. Let $\mathrm{Var}(y_0) = \sigma_0^2$.

$$\begin{aligned}&\mathrm{E}\left[y_0 - x_0'\hat{\beta} - \delta_*'(Y - X\hat{\beta})\right]^2 \\ &= \mathrm{E}\left[y_0 - \hat{E}(y_0|Y)\right]^2 + \mathrm{E}\left[\hat{E}(y_0|Y) - x_0'\hat{\beta} - \delta_*'(Y - X\hat{\beta})\right]^2 \\ &= \mathrm{E}\left[y_0 - \hat{E}(y_0|Y)\right]^2 + \mathrm{Var}\left[x_0'\hat{\beta} - \delta_*'X\hat{\beta}\right] \\ &= \sigma_0^2 - V_{y0}'V^-V_{y0} + (x_0' - \delta_*'X)(X'V^-X)^-(x_0 - X'\delta_*),\end{aligned}$$

or, writing the BLUP as $b'Y$, the prediction variance becomes

$$\mathrm{E}\left[y_0 - b'Y\right]^2 = \sigma_0^2 - 2b'V_{y0} + b'Vb.$$

EXAMPLE 12.2.4. *Prediction in Standard Linear Models.*
Our usual linear model situation is that the y_is have zero covariance and identical

variances. Thus, model (2) is satisfied with $V = \sigma^2 I$. A new observation y_0 would typically have zero covariance with the previous data, so $V_{y0} = 0$. It follows that $\delta_* = 0$ and the BLUP of y_0 is $x_0'\hat{\beta}$. This is just the BLUE of $\mathrm{E}(y_0) = x_0'\beta$.

EXAMPLE 12.2.5. *Spatial Data and Kriging.*
If y_i is an observation taken at some point in space, then x_i' typically contains the coordinates of the point. In dealing with spatial data, finding the BLUP is often called *Kriging*. The real challenge with spatial data is getting some idea of the covariance matrices V and V_{y0}. See Christensen (2001) for a more detailed discussion with additional references.

Exercise 12.1 Prove three facts about Kriging.

(a) If $b'Y$ is the BLUP of y_0 and if $x_{i1} = 1$, $i = 0, 1, \ldots, n$, then $b'J = 1$.

(b) If $(y_0, x_0') = (y_i, x_i')$ for some $i \geq 1$, then the BLUP of y_0 is just y_i.

(c) If V is nonsingular and $b'Y$ is the BLUP of y_0, then there exists a vector γ such that the following equation is satisfied:

$$\begin{bmatrix} V & X \\ X' & 0 \end{bmatrix} \begin{bmatrix} b \\ \gamma \end{bmatrix} = \begin{bmatrix} V_{y0} \\ x_0 \end{bmatrix}.$$

Typically, this equation will have a unique solution.

Hint: Recall that $x_0' = \rho' X$ and that $X(X'V^{-1}X)^- X'V^{-1}$ is a projection operator onto $C(X)$.

EXAMPLE 12.2.6. *Mixed Model Prediction.*
In the mixed model (12.1.1), we wish to find the BLUP of $\lambda'\gamma$ based on the data Y. The vector λ can be any known vector. Note that $\mathrm{E}(\lambda'\gamma) = 0$; so we can let $y_0 = \lambda'\gamma$ and $x_0' = 0$. From Section 1, $V = ZDZ' + R$ and

$$\begin{aligned} V_{y0} &= \mathrm{Cov}(Y, \lambda'\gamma) \\ &= \mathrm{Cov}(Z\gamma + e, \lambda'\gamma) \\ &= \mathrm{Cov}(Z\gamma, \lambda'\gamma) \\ &= ZD\lambda. \end{aligned}$$

The BLUP of $\lambda'\gamma$ is

$$\delta_*'(Y - X\hat{\beta}),$$

where $X\hat{\beta}$ is the BLUE of $X\beta$ and δ_* satisfies $(ZDZ' + R)\delta_* = ZD\lambda$. The matrices X, Z, and λ are all known. As in Kriging, the practical challenge is to get some idea of the covariance matrices. In this case, we need D and R. Estimates of D and R are one byproduct of variance component estimation. Estimation of variance components is discussed later in this chapter.

Exercise 12.2 If $\lambda_1'\beta$ is estimable, find the best linear unbiased predictor of

$\lambda_1'\beta + \lambda_2'\gamma$. For this problem, $b_0 + b'Y$ is unbiased if $\mathrm{E}(b_0 + b'Y) = \mathrm{E}(\lambda_1'\beta + \lambda_2'\gamma)$. The best predictor minimizes $\mathrm{E}[b_0 + b'Y - \lambda_1'\beta - \lambda_2'\gamma]^2$.

Exercise 12.3 Assuming the results of Exercise 6.3, show that the BLUP of the random vector $\Lambda'\gamma$ is $Q'(Y - X\hat{\beta})$, where $VQ = ZD\Lambda$.

12.3 Mixed Model Equations

We now develop the well-known *mixed model equations*. These equations are similar in spirit to normal equations; however, the mixed model equations simultaneously provide BLUEs and BLUPs.

Suppose that R is nonsingular. If γ were not random, the normal equations for model (12.1.1) would be

$$\begin{bmatrix} X' \\ Z' \end{bmatrix} R^{-1}[X, Z] \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X' \\ Z' \end{bmatrix} R^{-1}Y$$

or

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{bmatrix}.$$

For D nonsingular, the mixed model equations are defined as

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & D^{-1} + Z'R^{-1}Z \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{bmatrix}. \tag{1}$$

Theorem 12.3.1. If $[\hat{\beta}', \hat{\gamma}']$ is a solution to the mixed model equations, then $X\hat{\beta}$ is a BLUE of $X\beta$ and $\hat{\gamma}$ is a BLUP of γ.

PROOF. From Section 2.8, $X\hat{\beta}$ will be a BLUE of $X\beta$ if $\hat{\beta}$ is a solution to $X'V^{-1}X\beta = X'V^{-1}Y$. To use this equation we need a form for V^{-1} in terms of Z, D, and R:

$$V^{-1} = R^{-1} - R^{-1}Z\left[D^{-1} + Z'R^{-1}Z\right]^{-1} Z'R^{-1}.$$

This follows from Theorem B.56.

If $\hat{\beta}$ and $\hat{\gamma}$ are solutions, then the second row of the mixed model equations gives

$$Z'R^{-1}X\hat{\beta} + \left[D^{-1} + Z'R^{-1}Z\right]\hat{\gamma} = Z'R^{-1}Y$$

or

$$\hat{\gamma} = \left[D^{-1} + Z'R^{-1}Z\right]^{-1} Z'R^{-1}(Y - X\hat{\beta}). \tag{2}$$

The first row of the equations is

$$X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{\gamma} = X'R^{-1}Y.$$

Substituting for $\hat{\gamma}$ gives

$$X'R^{-1}X\hat{\beta}+X'R^{-1}Z\left[D^{-1}+Z'R^{-1}Z\right]^{-1}Z'R^{-1}(Y-X\hat{\beta})=X'R^{-1}Y$$

or

$$\begin{aligned}X'R^{-1}X\hat{\beta}-X'R^{-1}Z\left[D^{-1}+Z'R^{-1}Z\right]^{-1}Z'R^{-1}X\hat{\beta}&\\ =X'R^{-1}Y-X'R^{-1}Z\left[D^{-1}+Z'R^{-1}Z\right]^{-1}Z'R^{-1}Y,&\end{aligned}$$

which is $X'V^{-1}X\hat{\beta}=X'V^{-1}Y$. Thus, $\hat{\beta}$ is a generalized least squares solution and $X\hat{\beta}$ is a BLUE.

$\hat{\gamma}$ in (2) can be rewritten as

$$\begin{aligned}\hat{\gamma}&=\left(D\left[D^{-1}+Z'R^{-1}Z\right]-DZ'R^{-1}Z\right)\left[D^{-1}+Z'R^{-1}Z\right]^{-1}Z'R^{-1}(Y-X\hat{\beta})\\&=\left(DZ'R^{-1}-DZ'R^{-1}Z\left[D^{-1}+Z'R^{-1}Z\right]^{-1}Z'R^{-1}\right)(Y-X\hat{\beta})\\&=DZ'V^{-1}(Y-X\hat{\beta}),\end{aligned}$$

which is the BLUP of γ from Exercise 12.3, taking $Q=V^{-1}ZD$. □

The mixed model equations' primary usefulness is that they are relatively easy to solve. Finding the solution to

$$X'V^{-1}X\beta=X'V^{-1}Y$$

requires inversion of the $n\times n$ matrix V. The mixed model equations

$$\begin{bmatrix}X'R^{-1}X & X'R^{-1}Z\\ Z'R^{-1}X & D^{-1}+Z'R^{-1}Z\end{bmatrix}\begin{bmatrix}\beta\\ \gamma\end{bmatrix}=\begin{bmatrix}X'R^{-1}Y\\ Z'R^{-1}Y\end{bmatrix}$$

require computation of two inverses, D^{-1}, which is of the order of magnitude of the number of random effects, and R^{-1}, which generally is taken to be a diagonal matrix. If there are many observations relative to the number of random effects, it is easier to solve the mixed model equations. Of course, using Theorem B.56 to obtain V^{-1} for the generalized normal equations accomplishes the same thing.

An equivalent form of the mixed model equations that does not require D to be nonsingular is

$$\begin{bmatrix}X'R^{-1}X & X'R^{-1}ZD\\ Z'R^{-1}X & I+Z'R^{-1}ZD\end{bmatrix}\begin{bmatrix}\beta\\ \xi\end{bmatrix}=\begin{bmatrix}X'R^{-1}Y\\ Z'R^{-1}Y\end{bmatrix}. \tag{3}$$

Solutions $\hat{\beta}$, $\hat{\xi}$ have $X\hat{\beta}$ a BLUE of $X\beta$ and $D\hat{\xi}=\hat{\gamma}$ a BLUP of $\hat{\gamma}$.

Exercise 12.4 Even when D is singular, equation (3) has an advantage over equation (1) in that equation (3) does not require D^{-1}. Show that equations (1) and (3) are equivalent when D is nonsingular.

The mixed model equations can also be arrived at from a Bayesian argument. Consider the model

$$Y = X\beta + Z\gamma + e, \quad e \sim N(0, R),$$

and, as discussed in Section 2.9, incorporate partial prior information in the form

$$\gamma \sim N(0, D),$$

where D is again assumed to be nonsingular. A minor generalization of equation (2.9.3) allows the data Y to have an arbitrary nonsingular covariance matrix, so the Bayesian analysis can be obtained from fitting the generalized least squares model

$$\begin{bmatrix} Y \\ 0 \end{bmatrix} = \begin{bmatrix} X & Z \\ 0 & I \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} e \\ \tilde{e} \end{bmatrix}, \quad \begin{bmatrix} e \\ \tilde{e} \end{bmatrix} \sim N\left(\begin{bmatrix} 0_{n\times 1} \\ 0_{r\times 1} \end{bmatrix}, \begin{bmatrix} R & 0 \\ 0 & D \end{bmatrix} \right).$$

The generalized least squares estimates from this model will be the posterior means of β and γ, respectively. However, the generalized least squares estimates can be obtained from the corresponding normal equations, and the normal equations are the mixed model equations (1).

12.4 Variance Component Estimation: Maximum Likelihood

Assume that $Y \sim N(X\beta, V)$ and that V is nonsingular, so that the density of Y exists. In this chapter we write determinants as $|V| \equiv \det(V)$. The density of Y is

$$(2\pi)^{-n/2} |V|^{-1/2} \exp\left[-(Y - X\beta)'V^{-1}(Y - X\beta)/2 \right],$$

where $|V|$ denotes the determinant of V. (The earlier notation $\det(V)$ becomes awkward.) The log-likelihood is

$$L(X\beta, V) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|V|) - \frac{1}{2}(Y - X\beta)'V^{-1}(Y - X\beta).$$

For fixed V, this is maximized by minimizing $(Y - X\beta)'V^{-1}(Y - X\beta)$. It follows from Section 2.8 that, given the maximum likelihood estimate of V, the maximum likelihood estimate of $X\beta$ is obtained by treating the MLE of V as the true value of V and taking the usual generalized least squares estimator of $X\beta$.

Finding the MLE of V is rather more involved. In general, R could involve $n(n+1)/2$ distinct parameters, and D could involve another $q(q+1)/2$, where q is the number of columns in Z. There is little hope of estimating that many parameters from n observations. We need to assume some additional structure for the mixed model. Traditionally, variance component problems are analysis of variance problems with random rather than fixed effects. The levels of any particular effect are assumed to be independent with the same variance. Different effects are allowed different variances and are assumed independent. A natural generalization of model (12.1.1) is to partition Z into, say, r submatrices. Write $Z = [Z_1, Z_2, \ldots, Z_r]$ and $\gamma' = [\gamma_1', \gamma_2', \ldots, \gamma_r']$ with $\mathrm{Cov}(\gamma_i) = \sigma_i^2 I_{q(i)}$ and, for $i \neq j$, $\mathrm{Cov}(\gamma_i, \gamma_j) = 0$. The notation $q(i)$ is used to indicate the number of columns in Z_i. $I_{q(i)}$ is a $q(i) \times q(i)$ identity matrix. The covariance matrix of γ is

$$D = \text{Blk diag}[\sigma_i^2 I_{q(i)}].$$

As usual, assume $R = \sigma_0^2 I_n$. With these conventions we can write

$$V = \sigma_0^2 I_n + \sum_{i=1}^{r} \sigma_i^2 Z_i Z_i' = \sum_{i=0}^{r} \sigma_i^2 Z_i Z_i',$$

where we take $Z_0 \equiv I$. This model is used for all the variance component estimation methods that we will discuss.

To find the maximum likelihood estimates, take the partial derivatives of the log-likelihood and set them equal to zero. To find these equations, several results on matrix differentiation are needed.

Proposition 12.4.1.

(1) $\partial Ax \big/ \partial x = A$.
(2) $\partial x'Ax \big/ \partial x = 2x'A$.
(3) If A is a function of a scalar s,

$$\frac{\partial A^{-1}}{\partial s} = -A^{-1}\frac{\partial A}{\partial s}A^{-1}.$$

(4) For V as above,

$$\frac{\partial}{\partial \sigma_i^2}\log|V| = \mathrm{tr}\left[V^{-1}\frac{\partial V}{\partial \sigma_i^2}\right] = \mathrm{tr}\big(V^{-1}Z_iZ_i'\big)\,.$$

PROOF. (1), (2), and (3) are standard results. (4) can be found in Searle et al. (1992, p. 457). □

The partial derivatives of the log-likelihood are

$$\frac{\partial L}{\partial \beta} = -\beta'X'V^{-1}X + Y'V^{-1}X$$

$$\frac{\partial L}{\partial \sigma_i^2} = -\frac{1}{2}\mathrm{tr}\big(V^{-1}Z_iZ_i'\big) + \frac{1}{2}(Y - X\beta)'V^{-1}\frac{\partial V}{\partial \sigma_i^2}V^{-1}(Y - X\beta),$$

$i = 0, \ldots, r$. Setting the partial derivatives equal to zero gives

$$X'V^{-1}X\beta = X'V^{-1}Y$$

$$\mathrm{tr}\big(V^{-1}Z_iZ_i'\big) = (Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}(Y - X\beta),$$

$i = 0,\ldots,r$. Generally, some sort of iterative computation procedure is required to solve these equations. In particular, methods based on solving a system of linear equations similar to those discussed in Section 6 can be used.

It is interesting to note that solving the likelihood equations also gives method of moments (or estimating equation) estimates. The likelihood equations can be viewed as setting

$$X'V^{-1}Y = \mathrm{E}[X'V^{-1}Y]$$

and for $i = 0,\ldots,r$.

$$(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}(Y - X\beta) = \mathrm{E}[(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}(Y - X\beta)].$$

There are many questions about this technique. The MLEs may be solutions to the equations with $\sigma_i^2 > 0$ for all i, or they may not be solutions, but rather be on a boundary of the parameter space. There may be solutions other than the maximum. What are good computing techniques? These questions are beyond the scope of this book.

EXAMPLE 12.4.2. *Balanced One-Way ANOVA.*
Let $y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1,\ldots,t$, $j = 1,\ldots,N$, with the α_is independent $N(0,\sigma_1^2)$, the e_{ij}s independent $N(0,\sigma_0^2)$, and the α_is and e_{ij}s independent.

The matrix $[X,Z]$ for the general mixed model is just the model matrix from Chapter 4, where $X = J_n^1$, $r = 1$, and $Z_1 = Z = [X_1,\ldots,X_t]$. As in Chapter 11,

$$V = \sigma_0^2 I + \sigma_1^2 ZZ' = \sigma_0^2 I + N\sigma_1^2 M_Z,$$

where M_Z is the perpendicular projection matrix onto $C(Z)$. The inverse of V is easily seen to be

$$V^{-1} = \frac{1}{\sigma_0^2}\left[I - \frac{N\sigma_1^2}{\sigma_0^2 + N\sigma_1^2}M_Z\right],$$

cf. Proposition 12.11.1.

We can now find the estimates. It is easily seen that $C(VX) \subset C(X)$, so least squares estimates provide solutions to $X'V^{-1}X\beta = X'V^{-1}Y$. Simply put, $\hat{\mu} = \bar{y}_{\cdot\cdot}$.

For $i = 0$, the likelihood equation is

$$\mathrm{tr}(V^{-1}) = (Y - \hat{\mu}J)'V^{-2}(Y - \hat{\mu}J).$$

Observe that

$$\begin{aligned}
V^{-1}(Y - \hat{\mu}J) &= \frac{1}{\sigma_0^2}\left[I - \frac{N\sigma_1^2}{\sigma_0^2 + N\sigma_1^2}M_Z\right]\left(I - \frac{1}{n}J_n^n\right)Y \\
&= \frac{1}{\sigma_0^2}\left[\left(I - \frac{1}{n}J_n^n\right)Y - \frac{N\sigma_1^2}{\sigma_0^2 + N\sigma_1^2}\left(M_Z - \frac{1}{n}J_n^n\right)Y\right]
\end{aligned}$$

$$= \frac{1}{\sigma_0^2}\left[(I-M_Z)Y - \left(\frac{N\sigma_1^2}{\sigma_0^2+N\sigma_1^2}-1\right)\left(M_Z-\frac{1}{n}J_n^n\right)Y\right]$$
$$= \frac{1}{\sigma_0^2}(I-M_Z)Y - \frac{1}{\sigma_0^2+N\sigma_1^2}\left(M_Z-\frac{1}{n}J_n^n\right)Y.$$

Thus, evaluating $\mathrm{tr}(V^{-1})$ on the lefthand side and computing the squared length on the righthand side leads to the equation

$$\frac{Nt[\sigma_0^2+(N-1)\sigma_1^2]}{\sigma_0^2(\sigma_0^2+N\sigma_1^2)} = \frac{SSE}{\sigma_0^4} + \frac{SSTrts}{(\sigma_0^2+N\sigma_1^2)^2}. \tag{1}$$

For $i=1$, the likelihood equation is

$$\mathrm{tr}(V^{-1}ZZ') = (Y-\hat{\mu}J)'V^{-1}ZZ'V^{-1}(Y-\hat{\mu}J). \tag{2}$$

Using $ZZ' = NM_Z$ and the characterization of $V^{-1}(Y-\hat{\mu}J)$, the righthand side of (2) can be written

$$\begin{aligned}(Y-\hat{\mu}J)'V^{-1}ZZ'V^{-1}(Y-\hat{\mu}J) &= N(Y-\hat{\mu}J)'V^{-1}M_ZV^{-1}(Y-\hat{\mu}J)\\ &= \frac{N}{\left(\sigma_0^2+N\sigma_1^2\right)^2}Y'\left(M_Z-\frac{1}{n}J_n^n\right)Y\\ &= \frac{N\,SSTrts}{\left(\sigma_0^2+N\sigma_1^2\right)^2}.\end{aligned}$$

To evaluate the lefthand side of (2), note that

$$\begin{aligned}\mathrm{tr}(V^{-1}ZZ') &= \mathrm{tr}\left\{\frac{1}{\sigma_0^2}\left[I-\frac{N\sigma_1^2}{\sigma_0^2+N\sigma_1^2}M_Z\right]NM_Z\right\}\\ &= \frac{N}{\sigma_0^2}\mathrm{tr}\left[M_Z-\frac{N\sigma_1^2}{\sigma_0^2+N\sigma_1^2}M_Z\right]\\ &= \frac{N}{\sigma_0^2}\frac{\sigma_0^2}{\sigma_0^2+N\sigma_1^2}\mathrm{tr}(M_Z)\\ &= \frac{Nt}{\sigma_0^2+N\sigma_1^2}.\end{aligned}$$

Equation (2) becomes

$$\frac{Nt}{\sigma_0^2+N\sigma_1^2} = \frac{N\,SSTrts}{\left(\sigma_0^2+N\sigma_1^2\right)^2}$$

or

$$\sigma_0^2+N\sigma_1^2 = SSTrts\big/t. \tag{3}$$

Substituting equation (3) into equation (1) and multiplying through by σ_0^4 gives

$$\frac{Nt\sigma_0^2[\sigma_0^2 + (N-1)\sigma_1^2]}{(\sigma_0^2 + N\sigma_1^2)} = SSE + \frac{t\sigma_0^4}{(\sigma_0^2 + N\sigma_1^2)}$$

or

$$t(N-1)\sigma_0^2 = SSE.$$

The maximum likelihood estimates appear to be $\hat{\sigma}_0^2 = MSE$ and $\hat{\sigma}_1^2 = [(\hat{\sigma}_0^2 + N\hat{\sigma}_1^2) - \hat{\sigma}_0^2]/N = [SSTrts/t - MSE]/N$. However, this is true only if $SSTrts/t - MSE > 0$. Otherwise, the maximum is on a boundary, so $\hat{\sigma}_1^2 = 0$ and $\hat{\sigma}_0^2 = SSE/tN$.

The maximum likelihood procedure tends to ignore the fact that mean parameters are being fitted. In the one-way ANOVA example, the estimate of $\sigma_0^2 + N\sigma_1^2$ was $SSTrts/t$ instead of the unbiased estimate $MSTrts$. No correction was included for fitting the parameter μ. To avoid this, one can perform maximum likelihood on the residuals, the subject of Section 6.

12.5 Maximum Likelihood Estimation for Singular Normal Distributions

Maximum likelihood estimation involves maximizing the joint density of the observations over the possible values of the parameters. This assumes the existence of a density. A density cannot exist if the covariance matrix of the observations is singular, as is the case with residuals. We consider an approach that allows maximum likelihood estimation and show some uniqueness properties of the approach.

Suppose Y is a random vector in $\mathbf{R}^n$, $\mathrm{E}(Y) = \mu$, $\mathrm{Cov}(Y) = V$. If $r(V) = r < n$, then, as seen in Lemma 1.3.5, $\Pr[(Y-\mu) \in C(V)] = 1$ and $Y - \mu$ is restricted to an r-dimensional subspace of $\mathbf{R}^n$. It is this restriction of $Y - \mu$ to a subspace of $\mathbf{R}^n$ (with Lesbesgue measure zero) that causes the nonexistence of the density. We seek a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^r$ that will admit a density for the transformed random vector. The linear transformation should not lose any information and the MLEs should, in some sense, be the unique MLEs.

Suppose we pick an $n \times r$ matrix B with $C(B) = C(V)$. $B'Y$ together with a nonrandom function of Y can reconstruct Y with probability 1. Let M_V be the perpendicular projection operator onto $C(V)$, then $Y = M_V Y + (I - M_V)Y$. $M_V Y = B(B'B)^{-1}B'Y$ is a function of $B'Y$ while $(I - M_V)Y = (I - M_V)\mu$ with probability 1, because $\mathrm{Cov}[(I - M_V)Y] = (I - M_V)V(I - M_V) = 0$.

We would also like to see that $\mathrm{Cov}(B'Y) = B'VB$ is nonsingular, so that a density can exist. Since $C(B) = C(V)$ and V is symmetric, $V = BTB'$ for some symmetric matrix T. If T is nonsingular, then $B'VB = (B'B)T(B'B)$ is nonsingular, because both T and $B'B$ are nonsingular. We now show that T is nonsingular. Suppose T is singular. Then there exists $d \in \mathbf{R}^r$, $d \neq 0$, so that $Td = 0$. Since $r(B) = r$, there exists $b \neq 0$ such that $d = B'b$. Because $B'b = B'M_V b$, we can assume that $b \in C(V)$. Now, $Vb = BTB'b = BTd = 0$. However, for $b \in C(V)$, $Vb = 0$ can only happen if $b = 0$, which is a contradiction.

As mentioned earlier, there is little hope of estimating the entire matrix V. A more manageable problem is to assume that V is a function of a parameter vector θ. It should also be clear that in order to transform to a nonsingular random variable, we will need to know $C(V)$. This forces us to assume that $C(V(\theta))$ does not depend on θ.

Suppose now that $Y \sim N(\mu, V(\theta))$; then $B'Y \sim N(B'\mu, B'V(\theta)B)$. The density of $B'Y$ is

$$\begin{aligned}
&f(B'Y|\mu,\theta)\\
&= \frac{1}{(2\pi)^{\frac{r}{2}}}\frac{1}{|B'V(\theta)B|^{\frac{1}{2}}}\exp[-(B'Y-B'\mu)'[B'V(\theta)B]^{-1}(B'Y-B'\mu)/2]\\
&= (2\pi)^{-r/2}|B|^{-1}|V(\theta)|^{-1/2}\exp[-(Y-\mu)'B[B'V(\theta)B]^{-1}B'(Y-\mu)/2].
\end{aligned}$$

The MLEs are obtained by maximizing this with respect to μ and θ. A direct consequence of Proposition 12.5.1 below is that maximization of $f(B'Y)$ does not depend on the choice of B.

Proposition 12.5.1. If B and B_0 are two $n \times r$ matrices of rank r and $C(B) = C(B_0)$, then

(1) for some scalar k, $k|B'VB| = |B_0'VB_0|$;
(2) $B[B'VB]^{-1}B' = B_0[B_0'VB_0]^{-1}B_0'$ when the inverses exist.

PROOF. Since $C(B) = C(B_0)$ and both are full rank, $B_0 = BK$ for some nonsingular K.

(1) $|B_0'VB_0| = |K'B'VBK| = |K|^2|B'VB|$. Take $k = |K|^2$.
(2)
$$\begin{aligned}
B_0[B_0'VB_0]^{-1}B_0' &= BK[K'B'VBK]^{-1}K'B'\\
&= BKK^{-1}[B'VB]^{-1}(K')^{-1}K'B'\\
&= B[B'VB]^{-1}B'.
\end{aligned}$$
□

Corollary 12.5.2. $f(B'Y|\mu,\theta) = k^{-1/2}f(B_0'Y|\mu,\theta)$.

12.6 Variance Component Estimation: REML

Restricted (or residual) maximum likelihood (REML) estimation involves finding maximum likelihood estimates of variance components from the distribution of the residuals. This allows for estimation of the variance components without the complication of the fixed effects. An apparent problem is that of defining the residuals, since V is unknown. We show that any reasonable definition gives the same answers.

Consider the model of Section 4,

$$Y = X\beta + \xi, \quad \xi \sim N(0, V(\theta)),$$

$\theta = (\sigma_0^2, \ldots, \sigma_r^2)'$, $V(\theta) = \sum_{i=0}^{r} \sigma_i^2 Z_i Z_i'$. As discussed in Section 10.2, the only reasonable linear unbiased estimates of $X\beta$ are of the form AY, where A is some projection operator onto $C(X)$. The residuals can be defined as $(I-A)Y$. The distribution of the residuals is

$$(I-A)Y \sim N(0, (I-A)V(\theta)(I-A)').$$

$V(\theta)$ is assumed nonsingular in the mixed model, so $C((I-A)V(\theta)(I-A)') = C(I-A)$. Let $r(X) \equiv s$; so $r(I-A) = n-s$. For an $n \times s$ matrix B with $C(B) = C(I-A)$, the MLE of θ maximizes

$$f(B'(I-A)Y|\theta) = (2\pi)^{-(n-s)/2} |B'(I-A)V(\theta)(I-A)'B|^{-1/2} \\ \times \exp[-Y'(I-A)'B[B'(I-A)V(\theta)(I-A)'B]^{-1}B'(I-A)Y/2].$$

We will show that this depends on neither A nor B by showing that $C[(I-A)'B] = C(X)^{\perp}$ and appealing to Section 5.

Proposition 12.6.1. $\quad C((I-A)'B) = C(X)^{\perp}$.

PROOF. Clearly, $B'(I-A)X = 0$, so $C((I-A)'B) \subset C(X)^{\perp}$. The rank of $C(X)^{\perp}$ is $n-s$, so it is enough to show that the rank of $(I-A)'B$ is $n-s$. Since $(I-A)'B$ is an $n \times (n-s)$ matrix it is enough to show that for any $d \in \mathbf{R}^{n-s}$, $(I-A)'Bd = 0$ implies $d = 0$. Since $C(I-A) = C(B)$, $Bd = (I-A)c$ for some c. If $(I-A)'Bd = 0$, then $c(I-A)'Bd = d'B'Bd = 0$; so $Bd = 0$. Since B is an $n \times (n-s)$ matrix of full column rank, $d = 0$. $\square$

Frequently, REML is defined as maximum likelihood estimation from $B'Y$, where $r(B) = n-s$ and $B'X = 0$. This definition is equivalent to the one used here. Choose $A = M$. Then $I-M$ is the perpendicular projection operator onto $C(X)^{\perp}$. Since for any choice of B in this alternative definition, $r(B) = n-s$ and $B'X = 0$, we have $C(B) = C(X)^{\perp}$, and the B from this alternative definition also works in the original definition. The procedure presented here dictates doing maximum likelihood on $B'(I-M)Y = B'Y$. We will assume henceforth that $r(B) = n-s$ and $B'X = 0$.

As in Section 4, setting the partial derivatives to zero determines the likelihood equations

$$\mathrm{tr}[(B'VB)^{-1}B'Z_iZ_i'B] = Y'B(B'VB)^{-1}B'Z_iZ_i'B(B'VB)^{-1}B'Y, \tag{1}$$

$i = 0, 1, \ldots, r$. These can be written in a particularly nice form. To do this, we need some results on projection operators. Recall that if A is idempotent, then A is the projection operator onto $C(A)$ along $C(A')^{\perp}$. The rank of A is s, so $r(C(A')^{\perp})$ is $n-s$. Moreover, for any w, $w \in C(A')^{\perp}$ if and only if $Aw = 0$. To show that A is the projection operator onto $\mathscr{M}$ along $\mathscr{N}$ for two spaces $\mathscr{M}$ and $\mathscr{N}$, it is enough to show that 1) for $v \in \mathscr{M}$, $Av = v$; 2) for $w \in \mathscr{N}$, $Aw = 0$, and 3) $r(\mathscr{M}) + r(\mathscr{N}) = n$.

Lemma 12.6.2. Let $A_0 = X(X'V^{-1}X)^{-}X'V^{-1}$. Then $I - A_0$ is the projection operator onto $C(VB)$ along $C(X)$.

PROOF. Since $C(B) = C(X)^{\perp}$ and V is nonsingular, $r(VB) = n - s = r(I - A_0)$. Also, $(I - A_0)VB = VB - X(X'V^{-1}X)^{-}X'B = VB$, so $I - A_0$ is a projection onto $C(VB)$. It is along $C(X)$ because $r(X) = s$ and $(I - A_0)X = X - X = 0$. □

Lemma 12.6.3. $VB(B'VB)^{-1}B'$ is the projection operator onto $C(VB)$ along $C(X)$.

PROOF. $VB(B'VB)^{-1}B' = VB[(B'V)V^{-1}(VB)]^{-1}(B'V)V^{-1}$ is a projection onto $C(VB)$. Since $C(B) = C(X)^{\perp}$, $VB(B'VB)^{-1}B'X = 0$. Moreover, $r(VB) = n - s$ and $r(X) = s$. □

It follows from the lemmas that $(I - A_0) = VB(B'VB)^{-1}B'$ and that $V^{-1}(I - A_0) = B(B'VB)^{-1}B'$. Observing that $\mathrm{tr}[(B'VB)^{-1}B'Z_iZ_i'B] = \mathrm{tr}[B(B'VB)^{-1}B'Z_iZ_i']$, (1) can be rewritten as

$$\mathrm{tr}[V^{-1}(I - A_0)Z_iZ_i'] = Y'(I - A_0)'V^{-1}Z_iZ_i'V^{-1}(I - A_0)Y,$$

$i = 0, 1, \ldots, r$, where $(I - A_0)Y$ are the residuals from the BLUE of $X\beta$, i.e., $(I - A_0)Y = Y - X\hat{\beta}$.

As in Section 4, the REML equations can also be viewed as method of moments (estimating equation) methods because they involve setting a series of quadratic forms equal to their expectations. In particular, the REML equations involve setting

$$Y'(I - A_0)'V^{-1}Z_iZ_i'V^{-1}(I - A_0)Y = \mathrm{E}[Y'(I - A_0)'V^{-1}Z_iZ_i'V^{-1}(I - A_0)Y]$$

for $i = 0, 1, \ldots, r$. To see this observe that

$$\begin{aligned}
\mathrm{E}[Y'(I - A_0)'V^{-1}Z_iZ_i'V^{-1}(I - A_0)Y] &= \mathrm{tr}[(I - A_0)'V^{-1}Z_iZ_i'V^{-1}(I - A_0)V] \\
&= \mathrm{tr}[V^{-1}(I - A_0)V(I - A_0)'V^{-1}Z_iZ_i'] \\
&= \mathrm{tr}[V^{-1}(I - A_0)VV^{-1}(I - A_0)Z_iZ_i'] \\
&= \mathrm{tr}[V^{-1}(I - A_0)(I - A_0)Z_iZ_i'] \\
&= \mathrm{tr}[V^{-1}(I - A_0)Z_iZ_i'].
\end{aligned}$$

To solve the REML equations, it is useful to write

$$\begin{aligned}
\mathrm{tr}[V^{-1}(I - A_0)Z_iZ_i'] &= \mathrm{tr}[(I - A_0)'V^{-1}Z_iZ_i'V^{-1}(I - A_0)V] \\
&= \mathrm{tr}[V(I - A_0)'V^{-1}Z_iZ_i'V^{-1}(I - A_0)]
\end{aligned}$$

$$= \operatorname{tr}\left[\left(\sum_{j=0}^{r} \sigma_j^2 Z_j Z_j'\right)(I-A_0)'V^{-1}Z_iZ_i'V^{-1}(I-A_0)\right]$$
$$= \sum_{j=0}^{r} \sigma_j^2 \operatorname{tr}\left[Z_jZ_j'(I-A_0)'V^{-1}Z_iZ_i'V^{-1}(I-A_0)\right].$$

The equations for finding the REML estimates can now be written

$$\sum_{j=0}^{r} \sigma_j^2 \operatorname{tr}\left[Z_jZ_j'V^{-1}(I-A_0)Z_iZ_i'V^{-1}(I-A_0)\right]$$
$$= Y'(I-A_0)'V^{-1}Z_iZ_i'V^{-1}(I-A_0)Y,$$

$i = 0,\ldots,r$. Since V is unknown, typically an initial guess for V will be made and estimates of the σ_i^2 will be computed as the solution to the system of linear equations. These estimates of the variance components determine a new estimate of V that can be used to get updated values of the σ_i^2s. This iterative procedure is repeated until the σ_i^2s converge. Since the equations are linear, solutions are easy to find. As mentioned in Section 4, similar methods can be used to find unrestricted MLEs. In fact, this only involves removing the terms $(I-A_0)$ from the traces.

Exercise 12.5 Consider the model $Y = X\beta + e$, $e \sim N(0,\sigma^2 I)$. Show that the *MSE* is the REML estimate of σ^2.

12.7 Variance Component Estimation: MINQUE

MINQUE is short for *minimum norm quadratic unbiased (translation invariant) estimation*. Attention is restricted to estimates that are translation invariant unbiased quadratic forms in the observations.

Translation invariant means that a quadratic form, say $Y'BY$, has the property that $(Y+X\delta)'B(Y+X\delta) = Y'BY$ for all Y and δ. This simplifies to $-2Y'BX\delta = \delta'X'BX\delta$ for all Y and δ. In particular, $\delta'X'BX\delta = -2\mathrm{E}(Y)'BX\delta = -2\beta'X'BX\delta$ for all β and δ. It follows that $X'BX = 0$. Combining this with $\delta'X'BX\delta = -2Y'BX\delta$ implies that $Y'BX\delta = 0$ for all Y and δ. This can occur only if $BX = 0$. Conversely, if $BX = 0$, then $Y'BY$ is translation invariant.

As seen in Appendix B, $Y'BY = \operatorname{Vec}(B)'[Y \otimes Y]$, which is a linear function of $[Y \otimes Y]$. An alternative approach to variance component estimation involves fitting linear models to the vector $[Y \otimes Y]$, cf. Searle et al. (1992, Chapter 12) or Christensen (1993). In this section we use a somewhat less appealing but more elementary approach.

Minimum norm estimation considers unobservable "natural estimates" of the parameters $\sigma_0^2,\ldots,\sigma_r^2$, and seeks real estimates of $\sigma_0^2,\ldots,\sigma_r^2$ that are closest to the "natural" estimates. "Closest" could be defined by minimizing the distance between the vectors. Distance can be the usual Euclidean distance or some distance function that puts weights on the different dimensions of the vectors. A distance function is defined by a norm; hence the name "minimum norm."

Writing the model of Section 4 as

$$Y = X\beta + [Z_1, \ldots, Z_r] \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_r \end{bmatrix} + e,$$

with $e = (e_1, \ldots, e_n)'$ and $\gamma_j = (\gamma_{j1}, \ldots, \gamma_{jq(j)})'$. The "natural estimate" of σ_0^2 is $\sum_{i=1}^n e_i^2/n = e'e/n$, and a "natural estimate" of σ_j^2 is $\sum_{i=1}^{q(j)} \gamma_{ji}^2/q(j) = \gamma_j'\gamma_j/q(j)$. As a notational convenience, let $\gamma_0 \equiv e$.

Suppose now that we want to estimate $\sum_{i=0}^r p_i\sigma_i^2$ with an unbiased translation invariant quadratic form $Y'BY$. Because it is translation invariant, $Y'BY = (Y - X\beta)'B(Y - X\beta) = \sum_{i=0}^r \sum_{j=0}^r \gamma_i' Z_i' B Z_j \gamma_j$. To investigate unbiasedness, note that $\mathrm{E}(\gamma_i' Z_i' B Z_j \gamma_j) = \mathrm{E}[\mathrm{tr}(Z_i' B Z_j \gamma_j \gamma_i')] = \mathrm{tr}[\mathrm{E}(Z_i' B Z_j \gamma_j \gamma_i')] = \mathrm{tr}[Z_i' B Z_j \mathrm{Cov}(\gamma_j, \gamma_i)]$. If $i = j$, $\mathrm{Cov}(\gamma_i, \gamma_j) = \sigma_i^2 I_{q(i)}$. If $i \neq j$, $\mathrm{Cov}(\gamma_i, \gamma_j) = 0$; so $\mathrm{E}(Y'BY) = \sum_{i=0}^r \sigma_i^2 \mathrm{tr}(Z_i' B Z_i)$. Unbiasedness requires $\mathrm{tr}(Z_i' B Z_i) = p_i$. The quadratic form $Y'BY$ is using "estimates" $\gamma_i' Z_i' B Z_i \gamma_i / p_i$ for σ_i^2 and "estimates" $\gamma_i' Z_i' B Z_j \gamma_j$ for zero. Unfortunately, estimation procedures cannot depend on the unobservable γ_is.

It is reasonable to pick B to minimize

$$\sum_{i=0}^r w_i^2 [\gamma_i' Z_i' B Z_i \gamma_i - \gamma_i' \gamma_i (p_i/q(i))]^2 + \sum_{i \neq j} w_i w_j [\gamma_i' Z_i' B Z_j \gamma_j]^2.$$

This is a weighted distance between the "natural estimates" of the $p_i\sigma_i^2$s and the "estimates" of the $p_i\sigma_i^2$s implied by the use of $Y'BY$, plus a weighted distance between the covariances, which are zero, and the "estimates" of them from $Y'BY$. The weights, the w_is, are assumed to be fixed numbers supplied by the user. Unfortunately, without knowing the γ_is, this distance cannot be minimized. MINQUE instead minimizes

$$\sum_{i=0}^r \mathrm{tr}(w_i^2 [Z_i' B Z_i - (p_i/q(i)) I]^2) + \sum_{i \neq j} \mathrm{tr}(w_i w_j [Z_i' B Z_j][Z_j' B Z_i]).$$

This can be simplified somewhat. Since $\mathrm{tr}(Z_i' B Z_i) = p_i$,

$$\begin{aligned} \mathrm{tr}(w_i^2 [Z_i' B Z_i - (p_i/q(i)) I]^2) &= \mathrm{tr}(w_i^2 [(Z_i' B Z_i)^2 - 2 Z_i' B Z_i (p_i/q(i)) + (p_i/q(i))^2 I]) \\ &= \mathrm{tr}(w_i^2 (Z_i' B Z_i)^2) - 2 w_i^2 p_i^2/q(i) + w_i^2 p_i^2/q(i) \\ &= \mathrm{tr}(w_i^2 (Z_i' B Z_i)^2) - w_i^2 p_i^2/q(i). \end{aligned}$$

Only the term $\mathrm{tr}(w_i^2 (Z_i' B Z_i)^2)$ involves B. Therefore, the MINQUE estimate minimizes

$$\sum_{i=0}^{r}\sum_{j=0}^{r}\operatorname{tr}(w_i w_j [Z_i' B Z_j][Z_j' B Z_i])$$

with respect to B, subject to the conditions that $\operatorname{tr}(Z_i' B Z_i) = p_i$ and $BX = 0$. (B is assumed to be symmetric.)

Define $V_w = \sum_{i=0}^{r} w_i Z_i Z_i'$ and $A_w = X(X'V_w^{-1}X)^- X'V_w^{-1}$. With this notation the MINQUE is minimizing $\operatorname{tr}[(BV_w)^2]$ subject to $\operatorname{tr}(Z_i' B Z_i) = p_i$ and $BX = 0$. Rao (1971) has shown that the MINQUE estimate is $Y'B_0 Y$, where

$$B_0 = \sum_{i=0}^{r} \lambda_i (I - A_w)' V_w^{-1} Z_i Z_i' V_w^{-1} (I - A_w), \tag{1}$$

and where $\lambda_0, \ldots, \lambda_r$ are defined as solutions to the set of equations

$$\sum_{j=0}^{r} \lambda_j \operatorname{tr}[Z_j Z_j' V_w^{-1}(I - A_w) Z_i Z_i' V_w^{-1}(I - A_w)] = p_i, \tag{2}$$

$i = 0, \ldots, r$. Setting $Y'B_0 Y$ equal to $\sum_{j=0}^{r} p_j \hat{\sigma}_j^2$ and substituting for B_0 and p_i with (1) and (2), we can obtain estimates of the σ_i^2s.

$$\begin{aligned}
\sum_{i=0}^{r} \lambda_i Y'(I - A_w)' V_w^{-1} Z_i Z_i' V_w^{-1} (I - A_w) Y = Y' B_0 Y &= \sum_{i=0}^{r} p_i \hat{\sigma}_i^2 \\
&= \sum_{i=0}^{r} \hat{\sigma}_i^2 \sum_{j=0}^{r} \lambda_j \operatorname{tr}[Z_j Z_j' V_w^{-1}(I - A_w) Z_i Z_i' V_w^{-1}(I - A_w)].
\end{aligned}$$

Changing the order of summation on the last term, one sees that for any set of p_is, estimates can be obtained by solving

$$\begin{aligned}
\sum_{j=0}^{r} \hat{\sigma}_j^2 \operatorname{tr}[Z_j Z_j' V_w^{-1}(I - A_w) Z_i Z_i' V_w^{-1}(I - A_w)]& \\
= Y'(I - A_w)' V_w^{-1} Z_i Z_i' V_w^{-1} (I - A_w) Y&
\end{aligned}$$

for $i = 0, \ldots, r$.

These equations are precisely the REML equations with V_w and A_w substituted for V and A. MINQUE estimates with $V_w = V$ are REML estimates. The key difference in the methods is that V_w is a known matrix, whereas V is unknown.

If one finds the MINQUE estimates and then uses those estimates as weights in another round of MINQUE estimation, one has an iterative procedure that will eventually give REML estimates.

As mentioned, the choice of weights for MINQUE estimation is up to the user. A frequently used choice of weights is $w_i = 1$ for all i. This is the Euclidean norm. A very convenient set of weights is $w_0 = 1$, $w_i = 0$ for all $i \neq 0$. These are convenient in that with these weights, $V_w = I$ and $A_w = M$, the perpendicular projection matrix onto $C(X)$. The estimates obtained from these weights are often called

MINQUE(0) estimates and are sometimes called MIVQUE(0) estimates (see Section 8). Another possibility is to use an inexpensive estimation technique such as Henderson's Method 3 to obtain estimates $\hat{\sigma}_i^2$, $i = 0, \ldots, r$. The weights can be taken as $w_i = \hat{\sigma}_i^2$. The point of this adaptive method seems to be that the MINQUE estimates should then be very close to the REML estimates. In other words, with good initial weights, repeated iteration to the REML estimates should be unnecessary. One iteration may well do the job.

12.8 Variance Component Estimation: MIVQUE

MIVQUE is *minimum variance quadratic unbiased (translation invariant) estimation*. MIVQUE, like the maximum likelihood and REML procedures, and unlike MINQUE, assumes that Y has a normal distribution. We need to know the variance of a quadratic form.

Theorem 12.8.1. Suppose $Y \sim N(\mu, V)$ and V is nonsingular. Then $\mathrm{Var}(Y'BY) = 2\mathrm{tr}\left[(BV)^2\right] + 4\mu'BVB\mu$.

PROOF. See Searle (1971, Section 2.5). □

Since we are considering translation invariant quadratic forms $Y'BY$, we have $\mathrm{Var}(Y'BY) = 2\mathrm{tr}\left[(BV)^2\right]$.

If $Y'B_0Y$ is to be the minimum variance quadratic unbiased translation invariant estimate of $\sum_{i=0}^r p_i\sigma_i^2$, then $Y'B_0Y$ must minimize $\mathrm{tr}\left[(BV)^2\right]$ subject to $BX = 0$ and $\mathrm{tr}(Z_i'BZ_i) = p_i$. This is precisely the condition for getting a MINQUE estimate when $V_w = V$.

Unfortunately, it is virtually impossible to find MIVQUE estimates. In order to find B, one must minimize $\mathrm{tr}\left[(BV)^2\right]$. To do the minimization, one needs to know V. It is difficult for me to imagine a situation in which V would be known, yet not provide information about the variance components.

In practice, one can guess at V, find an appropriate MIVQUE matrix B based on the guess, obtain a new estimate of V, and repeat the procedure using the new V as the guess of V. This procedure should give something close to a MIVQUE, but it will not be exact. Because of the iterations involved, the estimates will not typically be quadratic unbiased translation invariant estimates. On the other hand, these iterated MIVQUE estimates can be viewed as iterated MINQUE estimates. Thus, they are precisely the REML estimates.

12.9 Variance Component Estimation: Henderson's Method 3

Long before the development of the other methods described, Henderson (1953) presented a way of obtaining unbiased method of moment estimates of variance components. Suppose we want to estimate σ_r^2. (The choice of σ_r^2 is for convenience only. The technique applies to estimating any of σ_1^2 to σ_r^2.) Let P be the perpendicular projection operator onto $C(X, Z_1, \ldots, Z_r)$ and P_0 the perpendicular projection operator onto $C(X, Z_1, \ldots, Z_{r-1})$. The expected value of $Y'(P - P_0)Y$ is

$$\mathrm{E}[Y'(P-P_0)Y] = \mathrm{tr}[(P-P_0)V] + \beta' X'(P-P_0)X\beta.$$

Since $C(X) \subset C(P_0) \subset C(P)$, $(P-P_0)X = 0$ and $\beta' X'(P-P_0)X\beta = 0$. Rewriting $(P-P_0)V$ gives

$$\begin{aligned}(P-P_0)V &= (P-P_0)\sigma_0^2 + \sum_{i=1}^{r} \sigma_i^2 (P-P_0)Z_i Z_i' \\ &= (P-P_0)\sigma_0^2 + (P-P_0)Z_r Z_r' \sigma_r^2,\end{aligned}$$

because $(P-P_0)Z_i = 0$ for $i = 1, \ldots, r-1$. It follows that

$$\begin{aligned}\mathrm{E}[Y'(P-P_0)Y] &= \mathrm{tr}\left[(P-P_0)\sigma_0^2 + (P-P_0)Z_r Z_r' \sigma_r^2\right] \\ &= \sigma_0^2 \mathrm{tr}(P-P_0) + \sigma_r^2 \mathrm{tr}\left[(P-P_0)Z_r Z_r'\right].\end{aligned} \tag{1}$$

A similar argument shows that

$$\mathrm{E}[Y'(I-P)Y] = \sigma_0^2 \mathrm{tr}(I-P).$$

An unbiased estimate of σ_0^2 is

$$\hat{\sigma}_0^2 = [Y'(I-P)Y] \big/ \mathrm{tr}(I-P). \tag{2}$$

From (1) and (2), an unbiased estimate of σ_r^2 is

$$\hat{\sigma}_r^2 = [Y'(P-P_0)Y - \hat{\sigma}_0^2 \mathrm{tr}(P-P_0)] \big/ \mathrm{tr}\left[(P-P_0)Z_r Z_r'\right], \tag{3}$$

provided that $P \neq P_0$.

Henderson's Method 3 has no known (to me) optimality properties. Henderson himself recommended the use of other techniques. Method 3's greatest advantage is that it is easy to compute. It uses standard techniques from fitting linear models by least squares, except that it requires the computation of $\mathrm{tr}[(P-P_0)Z_r Z_r']$. Even this can be computed using standard techniques. Note that

$$\mathrm{tr}\left[(P-P_0)Z_r Z_r'\right] = \mathrm{tr}\left[Z_r'(P-P_0)Z_r\right] = \mathrm{tr}\left[Z_r'(I-P_0)Z_r\right].$$

Write $Z_r = \left[Z_r^1, Z_r^2, \ldots, Z_r^{q(r)}\right]$, where each Z_r^j is a vector. By fitting the model

$$Z_r^j = X\beta + \sum_{i=1}^{r-1} Z_i\gamma_i + e$$

as if the γ_is are all fixed, we can obtain $(Z_r^j)'(I-P_0)(Z_r^j)$ as the sum of squares error, and thus

$$\mathrm{tr}\left[Z_r'(I-P_0)Z_r\right] = \sum_{j=1}^{q(r)} (Z_r^j)'(I-P_0)(Z_r^j).$$

In other words, all of the numbers required for estimating σ_r^2 can be obtained from a standard least squares computer program.

EXAMPLE 12.9.1. *Balanced One-Way ANOVA, continued from Example 12.4.2.* We relate the notation of this section to that of the earlier example:

$$P = M_Z,$$

$$P_0 = \frac{1}{n}J_n^n,$$

$$\mathrm{tr}(P-P_0) = t-1,$$

$$\mathrm{tr}(I-P) = t(N-1),$$

$$Y'(P-P_0)Y\big/\mathrm{tr}(P-P_0) = MSTrts,$$

$$Y'(I-P)Y\big/\mathrm{tr}(I-P) = MSE = \hat{\sigma}_0^2.$$

Recall that $ZZ' = NM_Z$, so

$$\begin{aligned}\mathrm{tr}\left[(P-P_0)ZZ'\right] &= \mathrm{tr}[(P-P_0)NP] \\ &= N\mathrm{tr}[(P-P_0)] \\ &= N(t-1).\end{aligned}$$

From (3) it is found that

$$\hat{\sigma}_1^2 = \frac{SSTrts - MSE(t-1)}{N(t-1)} = \frac{MSTrts - MSE}{N}.$$

Exercise 12.6 Consider the Method 3 estimates of σ_0^2 and σ_1^2 in Example 12.9.1.

(a) Show that these are also the REML estimates.

(b) Show that the vector $(Y'Y, Y'M_ZY, J'Y)'$ is a complete sufficient statistic for the balanced one-way random effects model.

(c) Show that the Method 3 estimates are minimum variance unbiased.

(d) Find the distributions of $SSTrts$ and SSE.

(e) Find a generalized likelihood ratio test of level α for $H_0 : \sigma_1^2 = 0$.

Hint: Use the concepts of Section 2.5 and the methods of Example 12.4.2.

For estimating more than one variance component, Method 3 is not, in general, well defined. Suppose that we desire to estimate both σ_r^2 and σ_{r-1}^2. To estimate σ_r^2, we proceed as before. To estimate σ_{r-1}^2, fit the model

$$Y = X\beta + \sum_{i=1}^{r-2} Z_i\gamma_i + e.$$

Let P_* be the perpendicular projection matrix onto $C(X, Z_1, \dots, Z_{r-2})$.

$$\begin{aligned}
\mathrm{E}[Y'(P_0 - P_*)Y] &= \mathrm{tr}[(P_0 - P_*)V] + \beta'X'(P_0 - P_*)X\beta \\
&= \mathrm{tr}\big[(P_0 - P_*)\sigma_0^2 + (P_0 - P_*)Z_{r-1}Z_{r-1}'\sigma_{r-1}^2 + (P_0 - P_*)Z_rZ_r'\sigma_r^2\big] \\
&= \sigma_0^2 t_0 + \sigma_{r-1}^2 t_{r-1} + \sigma_r^2 t_r,
\end{aligned}$$

where $t_0 = \mathrm{tr}(P_0 - P_*)$, $t_{r-1} = \mathrm{tr}[(P_0 - P_*)Z_{r-1}Z_{r-1}']$, and $t_r = \mathrm{tr}[(P_0 - P_*)Z_rZ_r']$. An unbiased estimate of σ_{r-1}^2 is

$$\hat{\sigma}_{r-1}^2 = \big[Y'(P_0 - P_*)Y - \hat{\sigma}_0^2 t_0 - \hat{\sigma}_r^2 t_r\big]\big/ t_{r-1}.$$

Note that t_0, t_{r-1}, and t_r can also be obtained from a least squares program.

The problem with this procedure is that the estimates depend on the order of estimation. One could equally well estimate σ_{r-1}^2 first with the technique as originally described, and use this second stage to estimate σ_r^2. Generally, the estimates of, say σ_r^2, will be different for the two methods of estimation. For *nested models*, however, only one order of estimation is possible. If $C(Z_{r-1}) \subset C(Z_r)$, we say that γ_r is nested within γ_{r-1}. We can estimate σ_r^2 first and then use it to estimate σ_{r-1}^2. The alternative order is not possible. If we desire to estimate σ_{r-1}^2 first, we require the perpendicular projection operator onto the orthogonal complement of $C(X, Z_1, \dots, Z_{r-2}, Z_r)$ with respect to $C(X, Z_1, \dots, Z_r)$. Because $C(Z_{r-1}) \subset C(Z_r)$, we have $C(X, Z_1, \dots, Z_{r-2}, Z_r) = C(X, Z_1, \dots, Z_r)$. The orthogonal complement is the zero space, and the projection matrix is the zero matrix.

For balanced ANOVA models, Henderson's Method 3 gives unique answers, because all of the effects are either orthogonal (e.g., main effects) or nested (e.g., a two-factor interaction is nested within both of the main effects). The definition of nested effects used here is somewhat nonstandard. To many people, interaction effects are not nested. As used here, interaction effects are nested in more than one other term.

Exercise 12.7 Data were generated according to the model

$$y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + e_{ijk},$$

$i = 1, 2$, $j = 1, 2, 3$, $k = 1, \dots, N_{ij}$, where $\mathrm{E}(\eta_j) = \mathrm{E}(\gamma_{ij}) = \mathrm{E}(e_{ijk}) = 0$, $\mathrm{Var}(e_{ijk}) = \sigma_0^2 = 64$, $\mathrm{Var}(\eta_j) = \sigma_1^2 = 784$, and $\mathrm{Var}(\gamma_{ij}) = \sigma_2^2 = 25$. All of the random effects were taken independently and normally distributed. The fixed effects were taken as $\mu+\alpha_1=200$ and $\mu+\alpha_2=150$. The data are

		j				j	
i	1	2	3	i	1	2	3
1	250	211	199	2	195	153	131
	262	198	184		187	150	133
	251	199	200		203	135	135
		198	187		192		
		184	184		209		
					184		

Estimate the variance components using MINQUE(0), MINQUE with all weights equal to 1, and Henderson's Method 3. Estimate the variance components using each of these three sets of estimates as starting values for REML and as starting values for maximum likelihood. For each method of estimation, find estimates of the fixed effects. Compare all the estimates with each other and with the true values. What tentative conclusions can you draw about the relative merits of the estimation procedures?

12.10 Exact *F* Tests for Variance Components

In this section we examine two procedures for testing whether a variance component is zero. The first test method is closely related to Henderson's estimation method.

12.10.1 Wald's Test

Seely and El-Bassiouni (1983) considered extensions of Wald's variance component test. They examined the mixed linear model

$$Y = X\beta + Z_1\gamma_1 + Z_2\gamma_2 + e. \tag{1}$$

Here, Z_1 and Z_2 are, respectively, $n \times q$ and $n \times s$ matrices of known quantities. γ_1, γ_2, and e are independent random vectors with

$$\gamma_1 \sim N(0,R), \quad \gamma_2 \sim N(0,\sigma_2^2 I_s), \quad e \sim N(0,\sigma_0^2 I_n).$$

The null hypothesis $H_0 : \sigma_2^2 = 0$ can be tested by using ordinary least squares calculations treating the γ_is as fixed effects. Let $SSE(1)$ be the sum of squares error from fitting model (1). The degrees of freedom error are $dfE(1)$. Also let $SSE(2)$ be the sum of squares error from the least squares fit of

$$Y = X\beta + Z_1\gamma_1 + e \tag{2}$$

with degrees of freedom error $dfE(2)$. The Wald test is simply based on the fact that under H_0

$$\frac{[SSE(2)-SSE(1)]/[dfE(2)-dfE(1)]}{SSE(1)/dfE(1)} \sim F(dfE(2)-dfE(1), dfE(1)).$$

Of course, if the two model matrix column spaces are the same, i.e., if $C(X,Z_1,Z_2) = C(X,Z_1)$, then there is no test because both the numerator sum of squares and degrees of freedom are zero.

We now verify the distribution given above. Note that under the mixed model,

$$\text{Cov}(Y) \equiv V = \sigma_0^2 I + \sigma_2^2 Z_2 Z_2' + Z_1 R Z_1'.$$

Let P_0 be the perpendicular projection operator onto the column space $C(X,Z_1)$, and let P be the perpendicular projection operator onto $C(X,Z_1,Z_2)$. It follows that $SSE(2) - SSE(1) = Y'(P-P_0)Y$, $dfE(2) - dfE(1) = r(P-P_0)$, $SSE(1) = Y'(I-P)Y$, and $dfE(1) = r(I-P)$. Note that $PZ_1 = Z_1$ and $PP_0 = P_0$. We need to show that $Y'(I-P)Y/\sigma_0^2 \sim \chi^2(r(I-P))$, that $Y'(I-P)Y$ and $Y'(P-P_0)Y$ are independent, and that, under H_0, $Y'(P-P_0)Y/\sigma_0^2 \sim \chi^2(r(P-P_0))$. Using results from Section 1.3, we need only show that $\sigma_0^{-4}(I-P)V(I-P) = \sigma_0^{-2}(I-P)$, that $(I-P)V(P-P_0) = 0$, and that, when $\sigma_2^2 = 0$, $\sigma_0^{-4}(P-P_0)V(P-P_0) = \sigma_0^{-2}(P-P_0)$. Verifying these results involves only straightforward linear algebra along with properties of projection operators. In general, the distribution of $Y'(P-P_0)Y$ seems to be intractable without the assumption that $\sigma_2^2 = 0$, but see Subsection 3.

To facilitate extensions of this test in the next subsection, we use a simple generalization of the Seely and El-Bassiouni results. The mixed model considered in (1) can also be written as

$$Y \sim N\left(X\beta, \sigma_0^2 I + \sigma_2^2 Z_2 Z_2' + Z_1 R Z_1'\right),$$

so the test applies for any data with a distribution of this form. If we let Σ be any known nonnegative definite matrix, the method also applies to

$$Y \sim N\left(X\beta, \sigma_0^2 \Sigma + \sigma_2^2 Z_2 Z_2' + Z_1 R Z_1'\right). \tag{3}$$

Simply write $\Sigma = QQ'$. Then there exists a matrix T such that $TQ = I$; so

$$TY \sim N\left(TX\beta, \sigma_0^2 I + \sigma_2^2 (TZ_2)(TZ_2)' + (TZ_1)R(TZ_1)'\right),$$

and the method applies to TY. Obviously, for Σ positive definite, $T = Q^{-1}$. The test based on (3) is simply the standard generalized least squares test of model (2) versus (1) when $\text{Cov}(Y) = \sigma_0^2 \Sigma$, see Section 3.8.

12.10.2 Öfversten's Second Method

Consider the special case of the model discussed in Subsection 1 with $R = \sigma_1^2 I_q$. Then $\text{Cov}(Y) \equiv V = \sigma_0^2 I + \sigma_2^2 Z_2 Z_2' + \sigma_1^2 Z_1 Z_1'$. The object of the second method is to obtain an exact F test for $H_0 : \sigma_1^2 = 0$. This is of primary interest when $C(X, Z_1) \subset C(X, Z_2)$. If $C(X, Z_1) \not\subset C(X, Z_2)$, a Wald test of $H_0 : \sigma_1^2 = 0$ is available by simply interchanging the roles of $Z_1\gamma_1$ and $Z_2\gamma_2$ in the previous section. If $C(X, Z_1) \subset C(X, Z_2)$, this interchange does not provide a test because then $C(X, Z_2) = C(X, Z_1, Z_2)$. As developed here, if Öfversten's second method provides a test, that test is valid regardless of the relationship of $C(X, Z_1)$ and $C(X, Z_2)$. At the end of the subsection, the test is presented for more general models.

Öfversten's (1993) second method as presented in Christensen (1996b) involves using an orthonormal basis. For example, use Gram–Schmidt on the columns of $[X, Z_2, Z_1, I_n]$ to obtain an orthonormal basis for $\mathbf{R}^n$, say $c_1, \ldots, c_n$. Write these as columns of a matrix $C = [c_1, \ldots, c_n]$. Partition C as $C = [C_1, C_2, C_3, C_4]$, where the columns of C_1 are an orthonormal basis for $C(X)$, the columns of C_2 are an orthonormal basis for the orthogonal complement of $C(X)$ with respect to $C(X, Z_2)$, the columns of C_3 are an orthonormal basis for the orthogonal complement of $C(X, Z_2)$ with respect to $C(X, Z_2, Z_1)$, and the columns of C_4 are an orthonormal basis for the orthogonal complement of $C(X, Z_2, Z_1)$. Note that if $C(X, Z_1) \subset C(X, Z_2)$, then C_3 is vacuous.

The basic idea of this method is to choose a matrix K so that the extended Wald's test for model (3) can be applied to $C_2'Y + KC_4'Y$. In executing the test of $H_0 : \sigma_1^2 = 0$, σ_1^2 plays the role assigned to σ_2^2 in (3), a function of σ_0^2 and σ_2^2 plays the role assigned to σ_0^2 in (3), and the role of R in (3) is vacuous. In particular, for some number λ and some matrix K, we want to have

$$C_2'Y + KC_4'Y \sim N\big(0, (\sigma_2^2 + \sigma_0^2/\lambda)C_2'Z_2Z_2'C_2 + \sigma_1^2 C_2'Z_1Z_1'C_2\big). \tag{4}$$

This is of the form (3). As shown in Exercise 12.8, $C_2'Z_2Z_2'C_2$ is a positive definite matrix, so the test follows immediately from generalized least squares. The test cannot be performed if $C_2'Z_1 = 0$, which occurs, for example, if $C(Z_1) \subset C(X)$. Note that $C_4C_4'Y$ is the vector of residuals from treating the random γ effects as fixed. Thus, in using $KC_4'Y = KC_4'C_4C_4'Y$ we are using some of the residual variability to construct the test.

To get the degrees of freedom for the test, we identify correspondences between (3) and (4). There are $r(C_2)$ "observations" available in (4). In (3), the numerator degrees of freedom for the test are $r(X, Z_1, Z_2) - r(X, Z_1)$. With mean zero in (4) there is no linear mean structure, i.e, nothing corresponding to X in (3), Z_1 from (3) is also vacuous in (4), and $C_2'Z_1$ is playing the role of Z_2 in (3). Thus the numerator degrees of freedom for the test are $r(C_2'Z_1)$ and the denominator degrees of freedom are $r(C_2) - r(C_2'Z_1)$. In model (4), $r(C_2) = r(X, Z_2) - r(X)$. If $C(Z_1) \subset C(X, Z_2)$, it is shown in Exercise 12.8 that $r(C_2'Z_1) = r(X, Z_1) - r(X)$ and the degrees of freedom for the test are $r(X, Z_1) - r(X)$ and $r(X, Z_1, Z_2) - r(X, Z_1)$, respectively.

Observe that

$$C_2'Y \sim N\left(0, \sigma_0^2 I + \sigma_2^2 C_2' Z_2 Z_2' C_2 + \sigma_1^2 C_2' Z_1 Z_1' C_2\right).$$

In many interesting cases, $C_2'Z_2Z_2'C_2 = \lambda I$; so an ordinary least squares Wald test can be applied immediately without any use of $C_4'Y$ as long as $C_2'Z_1 \neq 0$.

It is not difficult to see that in balanced ANOVA problems either $C_2'Z_2Z_2'C_2 = \lambda I$ when a standard balanced ANOVA test is available, or $C_2'Z_1 = 0$ when such a test is not available. For example, consider (1) as modeling a balanced two-way ANOVA $y_{ijk} = \mu + \gamma_{1i} + \gamma_{2j} + e_{ijk}$, $i = 1,\ldots,q$, $j = 1,\ldots,s$, $k = 1,\ldots,N$, with $X\beta$ being the vector $J\mu$. The γ_1 and γ_2 treatments are often said to be orthogonal. Letting M be the perpendicular projection operator onto $C(X)$, this orthogonality means precisely that $C[(I-M)Z_1]$ and $C[(I-M)Z_2]$ are orthogonal, i.e., $Z_2'(I-M)Z_1 = 0$, cf. Section 7.1. Now, by the definition of C_2, $C(C_2) = C[(I-M)Z_2]$, so $C_2'Z_1 = 0$ iff $Z_2'(I-M)Z_1 = 0$, which we know is true from orthogonality. Hence, no test of $H_0 : \sigma_1^2 = 0$ is available *from this method.*

Now consider a balanced nested model $y_{ijk} = \mu + \gamma_{1i} + \gamma_{2ij} + e_{ijk}$ with i, j, and k as above. In this model, $C(X, Z_1) \subset C(Z_2)$ and $\frac{1}{N}Z_2Z_2' \equiv P$ is the perpendicular projection operator onto $C(Z_2)$. Observing that $C(C_2) \subset C(Z_2)$ and using the orthonormality of the columns of C_2,

$$C_2'Z_2Z_2'C_2 = NC_2'PC_2 = NC_2'C_2 = NI.$$

Thus an ordinary least squares Wald test is available. Given that a Wald test simply compares models in the usual way, for $H_0 : \sigma_1^2 = 0$ this test is simply the standard balanced ANOVA test for no fixed γ_{1i} effects when γ_2 is random. Similar orthogonality and containment results hold in more general balanced ANOVAs. For the special case of $C(Z_1) \subset C(X, Z_2)$ with $C_2'Z_2Z_2'C_2 = \lambda I$, a general explicit form for the test statistic is given in (7).

In general, $C_2'Z_2Z_2'C_2 \neq \lambda I$, so the test requires $C_4'Y$. If $r(X, Z_2, Z_1) = t$,

$$C_4'Y \sim N\left(0, \sigma_0^2 I_{n-t}\right).$$

It is easy to see that $C_2'VC_4 = 0$, so $C_2'Y$ and $C_4'Y$ are independent. To obtain (4), simply pick K so that

$$KC_4'Y \sim N\left(0, \sigma_0^2\left[\lambda^{-1}C_2'Z_2Z_2'C_2 - I\right]\right).$$

Obviously one can do this provided that $\lambda^{-1}C_2'Z_2Z_2'C_2 - I$ is a nonnegative definite matrix. λ is chosen to ensure that the matrix is nonnegative definite. By the choice of C_2, $C_2'Z_2Z_2'C_2$ is a positive definite matrix and λ is taken as its smallest eigenvalue. Then if we use the eigenvalue, eigenvector decomposition $C_2'Z_2Z_2'C_2 = WD(\lambda_i)W'$ with W orthogonal,

$$\lambda^{-1}C_2'Z_2Z_2'C_2 - I = WD\left(\frac{\lambda_i}{\lambda} - 1\right)W',$$

which is clearly nonnegative definite.

Note that this development makes it obvious why λ needs to be the smallest eigenvalue. Actually, the test would still work if λ were chosen to be any positive number less than the smallest eigenvalue, but we want $KC_4'Y$ to increase the variability of $C_2'Y$ as little as possible, and this is accomplished by taking λ as large as possible. In particular, choosing λ as the smallest eigenvalue gives $KC_4'Y$ a singular covariance matrix and thus no variability in at least one direction. Other valid choices of λ can only increase variability.

Also note that $KC_4'Y = 0$ a.s. if the eigenvalues of $C_2'Z_2Z_2'C_2$ all happen to be the same. In this case, $\lambda^{-1}C_2'Z_2Z_2'C_2 - I = 0$ and $C_2'Z_2Z_2'C_2 = \lambda I$, so we get the simpler Wald test alluded to earlier.

The difficulty with the second method is that K is not unique, and typically the results of the test depend on the choice of K. In particular, K is a $w \times (n-t)$ matrix, where typically $w \equiv r(X,Z_2) - r(X) < (n-t)$, while $\lambda^{-1}C_2'Z_2Z_2'C_2 - I$ is a $w \times w$ matrix. Thus, we can take $K = \left[WD\left(\sqrt{\frac{\lambda_i}{\lambda} - 1}\right), 0\right]$ or $K = \left[0, WD\left(\sqrt{\frac{\lambda_i}{\lambda} - 1}\right)\right]$ or any number of other matrices. Modifying Öfversten, a reasonable procedure might be just to pick one of these convenient K matrices, but first randomly permute the rows of $C_4'Y$.

This method applies quite generally. A proof consists of observing that the test of $H_0 : \sigma_1^2 = 0$ remains valid when $X = [X_1, X_2]$, $\beta = (\beta_1', \beta_2')'$ with $\beta_2 \sim N(0, R_2)$, and β_2 independent of γ_1 and γ_2. This model allows for interaction between two random factors and arbitrary numbers of factors. The method will be most useful when $C(X,Z_1) \subset C(X,Z_2)$; if this is not the case, the simpler Wald test is available. Whenever $C_2'Z_1 = 0$, no test is available. For example, this will occur whenever $C(Z_1) \subset C(X)$, which is precisely what happens when one tries to test the variance component of a random main effect in a three-way analysis of variance with all interactions.

12.10.3 Comparison of Tests

When $C(Z_1) \not\subset C(X,Z_2)$, we have two tests of $H_0 : \sigma_1^2 = 0$ available. (See Lin and Harville (1991) and Christensen and Bedrick (1999) for some alternatives to Wald's test other than that just developed.) Let M, P_2, P, and P_0 be perpendicular projection matrices onto $C(X)$, $C(X,Z_2)$, $C(X,Z_2,Z_1)$, and $C(X,Z_1)$, respectively. The simple Wald test has the F statistic

$$F = \frac{Y'(P-P_2)Y/r(P-P_2)}{Y'(I-P)Y/r(I-P)}.$$

It can be of interest to examine the power (probability of rejecting the test) under some alternative to the null model, e.g., the model when the null hypothesis is false. The power of this test is quite complicated, but for given values of the parameters the power can be computed as in Davies (1980). Software is available through STATLIB. See also Christensen and Bedrick (1997).

Intuitively, the power depends in part (and only in part) on the degrees of freedom, $r(X,Z_2,Z_1) - r(X,Z_2)$, $n - r(X,Z_2,Z_1)$ and the ratio of the expected mean squares,

$$1 + \frac{\sigma_1^2}{\sigma_0^2} \frac{\operatorname{tr}[Z_1'(P-P_2)Z_1]}{r(X,Z_2,Z_1) - r(X,Z_2)}. \tag{5}$$

The basic idea behind F tests is that under the null hypothesis the test statistic is the ratio of two estimates of a common variance. Obviously, since the two are estimating the same thing under H_0, the ratio should be about 1. The F distribution quantifies the null variability about 1 for this ratio of estimates. If the numerator and denominator are actually estimates of very different things, the ratio should deviate substantially from the target value of 1. In fixed effects models, the power of an F test under the full model is simply a function of the ratio of expected values of the two estimates and the degrees of freedom of the estimates. In mixed models, the power is generally much more complicated, but the ratio of expected values can still provide some insight into the behavior of the tests. The ratio in (5) is strictly greater than 1 whenever the test exists and $\sigma_1^2 > 0$, thus indicating that larger values of the test statistic can be expected under the alternative. The power of the test should tend to increase as this ratio increases but in fact the power is quite complicated. Note that $Y'(P-P_2)Y = Y'C_3C_3'Y$ and $Y'(I-P)Y = Y'C_4C_4'Y$; so this test uses only $C_3'Y$ and $C_4'Y$.

The second test is based on $C_2'Y + KC_4'Y$. Again, exact powers can be computed as in Davies (1980). As shown in Exercise 12.8, the ratio of the expected mean squares for the second test is

$$1 + \frac{\sigma_1^2}{\sigma_2^2 + \sigma_0^2/\lambda} \frac{\operatorname{tr}\left[(C_2'Z_2Z_2'C_2)^{-1}C_2'Z_1Z_1'C_2\right]}{r(C_2'Z_1)}. \tag{6}$$

Again, this is strictly greater than 1 whenever the test exists and $\sigma_1^2 > 0$.

The degrees of freedom for the second test were given earlier. To compare the degrees of freedom for the two tests, observe that

$$C(X) \subset C(X,\{P_2 - M\}Z_1) = C(X,P_2Z_1) \subset C(X,Z_2) \subset C(X,Z_2,Z_1).$$

The degrees of freedom for the second test are, respectively, the ranks of the orthogonal complement of $C(X)$ with respect to $C(X,\{P_2 - M\}Z_1)$ and the orthogonal complement of $C(X,\{P_2 - M\}Z_1)$ with respect to $C(X,Z_2)$. (The first orthogonal complement is $C(C_2C_2'Z_1)$ with the same rank as $C_2'Z_1$, and the second orthogonal complement has rank $r(X,Z_2) - [r(X) + r(C_2'Z_1)]$.) The degrees of freedom for the simple Wald test are, respectively, the ranks of the orthogonal complement of $C(X,Z_2)$ with respect to $C(X,Z_2,Z_1)$ and the orthogonal complement of $C(X,Z_2,Z_1)$. In practice, the simple Wald test would typically have an advantage in having larger denominator degrees of freedom, but that could be outweighed by other factors in a given situation. We also see that, in some sense, the second test is being constructed inside $C(X,Z_2)$; it focuses on the overlap of $C(X,Z_2)$ and $C(Z_1)$. On the other hand, the simple Wald test is constructed from the overlap of $C(Z_1)$ with the orthogonal complement of $C(X,Z_2)$.

In the special case of $C(Z_1) \subset C(X,Z_2)$ with $C_2'Z_2Z_2'C_2 = \lambda I$, the second method gives the test statistic

$$F = \frac{Y'(P_0 - M)Y/r(P_0 - M)}{Y'(P_2 - P_0)Y/r(P_2 - P_0)}. \tag{7}$$

See Exercise 12.8 for a proof. For example, in a two-way ANOVA, X can indicate the grand mean and a fixed main effect, Z_1 can indicate the random main effect to be tested, and Z_2 can indicate the interaction. When the two-way is balanced, $C_2'Z_2Z_2'C_2 = \lambda I$ and we have the traditional test, i.e., the mean square for the random main effect divided by the mean square for interaction.

It should be noted that under $H_0 : \sigma_1^2 = 0$, $C_3'Y$ also has a $N(0, \sigma_0^2 I)$ distribution so it could also be used, along with $C_4'Y$, to adjust the distribution of $C_2'Y$ and still maintain a valid F test. However, this would be likely to have a deleterious effect on the power since then both the expected numerator mean square and the expected denominator mean square would involve positive multiples of σ_1^2 under the alternative.

The material in this section is closely related to Christensen (1996b). The near replicate lack of fit tests discussed in Subsection 6.6.2 can also be used to construct exact F tests for variance components. In fact, when used as a variance component test, Christensen's (1989) test is identical to Wald's test. See Christensen and Bedrick (1999) for an examination of these procedures.

Exercise 12.8

(a) Prove that $r(C_2'Z_1) = r(X,Z_1) - r(X)$ when $C(Z_1) \subset C(X,Z_2)$.
(b) Prove that $C_2'Z_2Z_2'C_2$ is positive definite.
(c) Prove (6).
(d) Prove (7).

Exercise 12.9 Use the data and model of Exercise 12.7 to test $H_0 : \sigma_1^2 = 0$ and $H_0 : \sigma_2^2 = 0$.

12.11 Recovery of Interblock Information in BIB Designs

Consider the analysis of a balanced incomplete block (BIB) design in which blocks are random effects. This analysis is known as the *recovery of interblock information*. The mixed model for BIBs was mentioned previously in Exercise 11.5. The analysis involves ideas from Chapters 9 through 11. BIB designs are discussed in Section 9.4, from which most of the current notation is taken. In estimating treatment effects, Theorem 10.4.5 is used. The model is a mixed model, so variance components are estimated using methods from this chapter. Finally, the variance structure breaks up into within cluster and between cluster components as in Chapter 11. In our analysis of the model, blocks are clusters, within cluster error is called *intrablock error*, and between cluster error is *interblock error*.

We begin by fixing notation and relating it back to Section 9.4. The model for a BIB is

$$y_{ij} = \mu + \beta_i + \tau_j + e_{ij},$$

$i = 1,\ldots,b$, with $j \in D_i$ or, equivalently, $j = 1,\ldots,t$, with $i \in A_i$. Here β and τ indicate block and treatment effects, respectively, D_i is the set of treatment indices for block i, and A_j is the set of indices for blocks that contain treatment j. The model is written using matrices as

$$Y = J\mu + X\beta + Z\tau + e,$$

where μ, β, and τ are the grand mean, block effects vector, and treatment effects vector, respectively. The matrix notation is a slight change from Section 9.4 in that J is no longer included in the X matrix. The matrix X is the matrix of indicators for the blocks and can be written

$$X = [x_{ij,m}], \qquad x_{ij,m} = \delta_{im},$$

where the columns of X are $m = 1,\ldots,b$ and the pair ij identifies a row of the matrix. Z is a matrix of indicators for the treatments and is defined as in Section 9.4, i.e., $Z = [z_{ij,r}]$ with $z_{ij,r} = \delta_{jr}$, $r = 1,\ldots,t$, and the pair ij denoting a row of the matrix. Recall two fundamental relations necessary for a BIB,

$$rt = bk$$

and

$$(t-1)\lambda = r(k-1),$$

where r is the number of replications for each treatment, k is the number of units in each block, and λ is the number of times any two treatments occur in the same block.

In the mixed model, β is a random vector with $\mathrm{E}(\beta) = 0$, $\mathrm{Cov}(\beta) = \sigma_B^2 I_b$, and $\mathrm{Cov}(\beta, e) = 0$. In a slight change of notation write $\mathrm{Cov}(e) = \sigma_e^2 I_n$, where $n = rt = bk$. Combining the random effects, write $\eta = X\beta + e$ and the model as

$$Y = Z\tau + \eta, \quad \mathrm{E}(\eta) = 0, \quad \mathrm{Cov}(\eta) = \sigma_e^2 I_n + \sigma_B^2 XX'. \tag{1}$$

Note that we have dropped the grand mean, thus removing the overparameterization associated with the treatment effects. In other words, we are using the model $y_{ij} = \tau_j + \eta_{ij}$, where $\eta_{ij} \equiv \beta_j + e_{ij}$ is the random error term.

As in Chapter 11, write $\sigma^2 = \sigma_e^2 + \sigma_B^2$ and $\rho = \sigma_B^2 / \left(\sigma_e^2 + \sigma_B^2\right)$. It follows that $\sigma_e^2 = \sigma^2(1-\rho)$ and $\sigma_B^2 = \sigma^2\rho$. A term that frequently appears in the analysis is the interblock (between cluster) error term,

$$\sigma^2\left[(1-\rho)+k\rho\right]=\sigma_e^2+k\sigma_B^2.$$

With the notation given earlier, write

$$\mathrm{Cov}(\eta)=\sigma^2 V,$$

where, again as in Chapter 11,

$$\begin{aligned} V &= \left[(1-\rho)I+\rho XX'\right] \\ &= \left[(1-\rho)I+k\rho M\right] \end{aligned}$$

and M is the perpendicular projection operator onto $C(X)$.

12.11.1 Estimation

In this subsection we derive the BLUE of τ. From Section 2.7,

$$\begin{aligned} Z\hat{\tau} &= AY \\ &= Z\left(Z'V^{-1}Z\right)^{-1}Z'V^{-1}Y. \end{aligned}$$

Note that finding $\hat{\tau}$ is essentially equivalent to finding the oblique projection operator A. Given $\hat{\tau}$ we can easily find $Z\hat{\tau}$; thus we know AY. With AY known for any vector Y, the matrix A is completely characterized. Finding $\hat{\tau}=\left(Z'V^{-1}Z\right)^{-1}Z'V^{-1}Y$ requires computation of both V^{-1} and $\left(Z'V^{-1}Z\right)^{-1}$. These computations are facilitated by the following result.

Proposition 12.11.1. Let P be a projection operator (idempotent), and let a and b be real numbers. Then

$$[aI+bP]^{-1}=\frac{1}{a}\left[I-\frac{b}{a+b}P\right].$$

PROOF.

$$\frac{1}{a}\left[I-\frac{b}{a+b}P\right][aI+bP]=\frac{1}{a}\left[aI+bP-\frac{ab}{a+b}P-\frac{b^2}{a+b}P\right]=I. \qquad \square$$

Using this result, we obtain two forms for V^{-1}:

$$V^{-1}=\frac{1}{1-\rho}\left[I-\frac{k\rho}{[(1-\rho)+k\rho]}M\right]=\frac{1}{1-\rho}\left[(I-M)+\frac{1-\rho}{[(1-\rho)+k\rho]}M\right]. \quad (2)$$

Both forms will be used frequently.

We now compute $Z'V^{-1}Z$ and $\left(Z'V^{-1}Z\right)^{-1}$. First note that the fundamental relation $(t-1)\lambda = r(k-1)$ implies

$$\lambda t = rk - (r-\lambda).$$

Using this equality, the characterizations of $Z'Z$ and $Z'MZ$ given in Section 9.4, the first form for V^{-1} in equation (2), and writing $P_t = \frac{1}{t}J_t^t$, we get

$$\begin{aligned}
(Z'V^{-1}Z) &= \frac{1}{1-\rho}\left[Z'Z - \frac{k\rho}{[(1-\rho)+k\rho]}Z'MZ\right] \\
&= \frac{1}{1-\rho}\left[rI - \frac{\rho}{[(1-\rho)+k\rho]}\{(r-\lambda)I + \lambda t P_t\}\right] \\
&= \frac{1}{1-\rho}\left[\frac{r(1-\rho)}{[(1-\rho)+k\rho]}I + \frac{\lambda t\rho}{[(1-\rho)+k\rho]}(I-P_t)\right] \\
&= \frac{1}{(1-\rho)[(1-\rho)+k\rho]}\left[r(1-\rho)I + \lambda t\rho(I-P_t)\right].
\end{aligned}$$

From Proposition 12.11.1,

$$\begin{aligned}
(Z'V^{-1}Z)^{-1} &= \frac{[(1-\rho)+k\rho]}{r}\left[I - \frac{\lambda t\rho}{r(1-\rho)+\lambda t\rho}(I-P_t)\right] \qquad (3)\\
&= \frac{[(1-\rho)+k\rho]}{r}\left[\frac{r(1-\rho)}{r(1-\rho)+\lambda t\rho}I + \frac{\lambda t\rho}{r(1-\rho)+\lambda t\rho}P_t\right].
\end{aligned}$$

Rather than computing $\hat{\tau} = (Z'V^{-1}Z)^{-1}Z'V^{-1}Y$ directly, it is convenient to decompose $Z'V^{-1}Y$ into intrablock and interblock components. The intrablock component is related to the fixed block effect analysis. The interblock component is what is left. The fixed block effect analysis is based on

$$Q = (Q_1, \ldots, Q_t)',$$

where $Q \equiv Z'(I-M)Y$, cf. Section 9.4. Similarly, define

$$W = (W_1, \ldots, W_t)',$$

where $W \equiv Z'MY$. M is the perpendicular projection operator for the one-way ANOVA in blocks (ignoring treatments) so

$$MY = [t_{ij}], \qquad t_{ij} = \bar{y}_{i\cdot},$$

with

$$\bar{y}_{i\cdot} \equiv \frac{1}{k}\sum_{j\in D_i} y_{ij}.$$

Z is a matrix of treatment indicators, so $Z'MY$ yields

$$W_j = \sum_{i \in A_j} \bar{y}_{i\cdot}\,.$$

In Section 9.4, the W_js were computed in the process of computing the Q_js. In particular,

$$Q_j = \sum_{i \in A_j} (y_{ij} - \bar{y}_{i\cdot}) = \sum_{i \in A_j} y_{ij} - W_j,$$

as in the first table in the continuation of Example 9.4.1. In computing $\hat{\tau}$, we will also have occasion to use

$$\overline{W}_{\cdot} \equiv \frac{1}{t} \sum_{j=1}^{t} W_j$$

and the fact that

$$\overline{Q}_{\cdot} \equiv \frac{1}{t} \sum_{j=1}^{t} Q_j = \frac{1}{t} J_t' Z'(I-M)Y = \frac{1}{t} J_n'(I-M)Y = 0.$$

Using the second form of V^{-1} given in (2),

$$\begin{aligned} Z'V^{-1}Y &= \frac{1}{1-\rho} Z'(I-M)Y + \frac{1}{(1-\rho)+k\rho} Z'MY \qquad (4) \\ &= \frac{1}{1-\rho} Q + \frac{1}{(1-\rho)+k\rho} W. \end{aligned}$$

Finally, using (3) and (4),

$$\begin{aligned} \hat{\tau} &= (Z'V^{-1}Z)^{-1} Z'V^{-1}Y \\ &= \frac{[(1-\rho)+k\rho]}{r} \left[\frac{r(1-\rho)}{r(1-\rho)+\lambda t\rho} I + \frac{\lambda t\rho}{r(1-\rho)+\lambda t\rho} P_t \right] Z'V^{-1}Y \\ &= \frac{[(1-\rho)+k\rho](1-\rho)}{r(1-\rho)+\lambda t\rho} Z'V^{-1}Y + \frac{[(1-\rho)+k\rho]\lambda t\rho}{r[r(1-\rho)+\lambda t\rho]} P_t Z'V^{-1}Y \\ &= \frac{[(1-\rho)+k\rho]}{r(1-\rho)+\lambda t\rho} Q + \frac{(1-\rho)}{r(1-\rho)+\lambda t\rho} W + \frac{\lambda t\rho \overline{W}_{\cdot}}{r[r(1-\rho)+\lambda t\rho]} J_t. \end{aligned}$$

The last equality comes from $P_t Q = \overline{Q}_{\cdot} J_t = 0$ and $P_t W = \overline{W}_{\cdot} J_t$. In particular, an individual component of $\hat{\tau}$ is

$$\hat{\tau}_j = \frac{[(1-\rho)+k\rho]}{r(1-\rho)+\lambda t\rho} Q_j + \frac{(1-\rho)}{r(1-\rho)+\lambda t\rho} W_j + \frac{\lambda t\rho \overline{W}_{\cdot}}{r[r(1-\rho)+\lambda t\rho]}.$$

For purposes of comparing treatments, the term involving $\overline{W}_{\cdot}$, which is constant, can be dropped. Finally, the projection operator is characterized by

$$AY = Z\hat{\tau} = [t_{ij}], \qquad t_{ij} = \hat{\tau}_j\,.$$

There are three additional aspects of the analysis to consider. First, we need to consider testing the hypothesis $\tau_1 = \cdots = \tau_t$. Second, we need to examine contrasts. Third, we need to deal with the fact that our estimate of $\hat{\tau}$ is useless. The estimate depends on $\rho = \sigma_B^2/(\sigma_e^2 + \sigma_B^2)$. This is an unknown parameter. Writing $\hat{\tau} = (\sigma^2/\sigma^2)\hat{\tau}$, and using $\sigma^2(1-\rho) = \sigma_e^2$ and $\sigma^2\rho = \sigma_B^2$, gives

$$\hat{\tau} = \frac{\sigma_e^2 + k\sigma_B^2}{r\sigma_e^2 + \lambda t\sigma_B^2} Q + \frac{\sigma_e^2}{r\sigma_e^2 + \lambda t\sigma_B^2} W + \frac{\lambda t\sigma_B^2 \overline{W}.}{r\sigma_e^2 + \lambda t\sigma_B^2} J_t. \tag{5}$$

Model (1) is a mixed model, so the methods of this chapter can be used to estimate σ_e^2 and σ_B^2. The variance estimates can be substituted into (5) to give a usable estimate of τ. Traditionally, Henderson's Method 3 has been used to obtain variance estimates. The use of Henderson's Method 3 will be discussed in detail later. Tests of models and examination of contrasts will be discussed as if σ_e^2 and σ_B^2 (hence σ^2 and ρ) were known. A discussion in which only ρ is assumed known is also given. Throughout we assume that $\eta \equiv X\beta + e \sim N(0, \sigma^2 V)$.

12.11.2 Model Testing

We desire to test model (1) against the reduced model

$$Y = J\mu + \eta. \tag{6}$$

In particular, we will show that an α level test rejects H_0 if

$$\frac{r\sigma_e^2 + \lambda t\sigma_B^2}{\sigma_e^2\left[\sigma_e^2 + k\sigma_B^2\right]} \sum_{i=1}^{t} (\hat{\tau}_j - \tilde{\tau}.) > \chi^2(1-\alpha, t-1), \tag{7}$$

where $\tilde{\tau}. = \sum_{j=1}^{t} \hat{\tau}_j/t$ is the mean of the $\hat{\tau}_j$s. The remainder of this subsection is devoted to showing that this is the appropriate test.

We begin by finding the BLUE of $J\mu$ from model (6). The BLUE is $J\hat{\mu} = A_0 Y = J\left(J'V^{-1}J\right)^{-1} J'V^{-1}Y$. However, we can apply Theorem 10.4.5 to see that the simple least squares estimate $J\hat{\mu}$ with $\hat{\mu} = \sum_{ij} y_{ij}/rt$ is the BLUE. To use Theorem 10.4.5, we need to show that $C(VJ) \subset C(J)$. Because $J \in C(X)$, $VJ = [(1-\rho)I + k\rho M]J = (1-\rho)J + k\rho J = [(1-\rho) + k\rho]J \in C(J)$.

From Corollary 3.8.3,

$$Y'(A - A_0)'V^{-1}(A - A_0)Y/\sigma^2 \sim \chi^2(t-1, 0)$$

if and only if model (6) is true. We wish to show that the test statistic is identical to that used in (7). Our argument involves five algebraic identities. First,

$$\sum_{j=1}^{t} (\hat{\tau}_j - \tilde{\tau}.)^2 = \hat{\tau}'(I - P_t)\hat{\tau}.$$

Second, we show that $\hat{\mu} = \tilde{\tau}_{\cdot}$. Using the second form for V^{-1} in (2) gives $V^{-1}J = [(1-\rho)+k\rho]^{-1}J$; also recall that $J \in C(Z)$ so $AJ = J$. These equalities lead to the result.

$$\begin{aligned}
\tilde{\tau}_{\cdot} &= \frac{1}{t}J_t'\hat{\tau} = \frac{1}{rt}J'Z\hat{\tau} = \frac{1}{rt}J'AY \\
&= \frac{(1-\rho)+k\rho}{rt}J'V^{-1}AY \\
&= \frac{(1-\rho)+k\rho}{rt}J'A'V^{-1}Y \\
&= \frac{(1-\rho)+k\rho}{rt}J'V^{-1}Y \\
&= \frac{1}{rt}J'Y = \hat{\mu}.
\end{aligned}$$

Third,

$$\begin{aligned}
\hat{\tau}'(I-P_t)\hat{\tau} &= \hat{\tau}'\hat{\tau} - t\left(\frac{1}{t}\hat{\tau}'J_t'\right)^2 \\
&= \hat{\tau}'\hat{\tau} - t\hat{\mu}^2 \\
&= [\hat{\tau} - J_t\hat{\mu}]'[\hat{\tau} - J_t\hat{\mu}].
\end{aligned}$$

Fourth, recall from Section 9.4 that

$$Z'(I-M)Z = \frac{\lambda t}{k}(I-P_t),$$

and, finally, from Section 9.4, and because $r-\lambda = rk-\lambda t$,

$$\begin{aligned}
Z'MZ &= \frac{1}{k}[(r-\lambda)I + \lambda t P_t] \\
&= \frac{1}{k}[rkI - \lambda t(I-P_t)].
\end{aligned}$$

Using the second form of V^{-1} in (2),

$$\begin{aligned}
&Y'(A-A_0)'V^{-1}(A-A_0)Y \\
&= [Z\hat{\tau} - J\hat{\mu}]'V^{-1}[Z\hat{\tau} - J\hat{\mu}] \\
&= \frac{1}{1-\rho}[Z\hat{\tau} - J\hat{\mu}]'(I-M)[Z\hat{\tau} - J\hat{\mu}] + \frac{1}{(1-\rho)+k\rho}[Z\hat{\tau} - J\hat{\mu}]'M[Z\hat{\tau} - J\hat{\mu}] \\
&= \frac{1}{1-\rho}\hat{\tau}'Z'(I-M)Z\hat{\tau} + \frac{1}{(1-\rho)+k\rho}[Z\hat{\tau} - ZJ_t\hat{\mu}]'M[Z\hat{\tau} - ZJ_t\hat{\mu}] \\
&= \frac{1}{1-\rho}\frac{\lambda t}{k}\hat{\tau}'(I-P_t)\hat{\tau} + \frac{1}{(1-\rho)+k\rho}[\hat{\tau} - J_t\hat{\mu}]'Z'MZ[\hat{\tau} - J_t\hat{\mu}]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{1-\rho}\,\frac{\lambda t}{k}\,\hat{\tau}'(I-P_t)\hat{\tau} + \frac{1}{(1-\rho)+k\rho}[\hat{\tau}-J_t\hat{\mu}]'\left\{\frac{1}{k}\left[rkI-\lambda t\,(I-P_t)\right]\right\}[\hat{\tau}-J_t\hat{\mu}] \\
&= \frac{\lambda t}{k}\,\hat{\tau}'(I-P_t)\hat{\tau}\left\{\frac{1}{1-\rho} - \frac{1}{(1-\rho)+k\rho}\right\} + \frac{r}{(1-\rho)+k\rho}[\hat{\tau}-J_t\hat{\mu}]'[\hat{\tau}-J_t\hat{\mu}] \\
&= \frac{k\rho}{[(1-\rho)+k\rho]\,(1-\rho)}\,\frac{\lambda t}{k}\,\hat{\tau}'(I-P_t)\hat{\tau} + \frac{r}{(1-\rho)+k\rho}\hat{\tau}'(I-P_t)\hat{\tau} \\
&= \frac{r(1-\rho)+\lambda t\rho}{(1-\rho)\,[(1-\rho)+k\rho]}\,\hat{\tau}'(I-P_t)\hat{\tau} \\
&= \frac{r\sigma^2(1-\rho)+\lambda t\sigma^2\rho}{\sigma^2(1-\rho)\,[(1-\rho)+k\rho]}\,\hat{\tau}'(I-P_t)\hat{\tau} \\
&= \frac{r\sigma_e^2+\lambda t\sigma_B^2}{\sigma_e^2\,[(1-\rho)+k\rho]}\,\hat{\tau}'(I-P_t)\hat{\tau}.
\end{aligned}
$$

Dividing by σ^2 gives

$$\frac{Y'\,(A-A_0)'\,V^{-1}\,(A-A_0)\,Y}{\sigma^2} = \frac{r\sigma_e^2+\lambda t\sigma_B^2}{\sigma_e^2\left[\sigma_e^2+k\sigma_B^2\right]}\,\hat{\tau}'(I-P_t)\hat{\tau},$$

which is the test statistic in (7). In practice, estimates of σ_e^2 and σ_B^2 are used to compute both the multiplier and $\hat{\tau}$. The substitution is then ignored and the χ^2 test is conducted as if σ_e^2 and σ_B^2 were known.

12.11.3 Contrasts

Model (1) is a one-way ANOVA model with an unusual covariance structure. However, estimable functions do not depend on the covariance matrix, so contrasts are estimable. This is true regardless of whether μ is included as a parameter in the model. A contrast is a linear parametric function $\xi'\tau$ with $\xi'J_t = 0$. The estimate is $\xi'\hat{\tau} = \sum_{j=1}^{t}\xi_j'\hat{\tau}_j$, where $\hat{\tau}$ has already been characterized.

We need to compute the variance of the estimate. Recall that with $\xi'J_t = 0$ we have $\xi'P_t = 0$. Using the second form in (3),

$$
\begin{aligned}
\mathrm{Var}\left(\xi'\hat{\tau}\right) &= \xi'\mathrm{Cov}(\hat{\tau})\xi \\
&= \sigma^2\xi'\left(Z'V^{-1}Z\right)^{-1}\xi \\
&= \sigma^2\frac{[(1-\rho)+k\rho]\,(1-\rho)}{r(1-\rho)+\lambda t\rho}\xi'\xi + \sigma^2\frac{[(1-\rho)+k\rho]\,\lambda t\rho}{r\,[r(1-\rho)+\lambda t\rho]}\xi'P_t\xi \\
&= \frac{\left[\sigma^2(1-\rho)+k\sigma^2\rho\right]\sigma^2(1-\rho)}{r\sigma^2(1-\rho)+\lambda t\sigma^2\rho}\xi'\xi \\
&= \frac{\left[\sigma_e^2+k\sigma_B^2\right]\sigma_e^2}{r\sigma_e^2+\lambda t\sigma_B^2}\xi'\xi.
\end{aligned}
$$

Note that the variance can also be written as

$$\mathrm{Var}(\xi'\hat{\tau}) = \sigma_e^2 \frac{[(1-\rho)+k\rho]}{r(1-\rho)+\lambda t\rho}\xi'\xi;$$

this second form will be used in the next subsection. Under normality,

$$\xi'\hat{\tau} \sim N\left(\xi'\tau, \frac{\left[\sigma_e^2 + k\sigma_B^2\right]\sigma_e^2}{r\sigma_e^2+\lambda t\sigma_B^2}\xi'\xi\right). \tag{8}$$

In practice, estimates of σ_e^2 and σ_B^2 are substituted to find $\hat{\tau}$ and the estimated variance. Tests and confidence intervals are conducted using the distribution (8), ignoring the fact that estimates have been substituted for σ_e^2 and σ_B^2.

12.11.4 Alternative Inferential Procedures

Traditionally, statistical inferences have been conducted using the distributions in (7) and (8). These are based on the incorrect assumption that both σ_e^2 and σ_B^2 are known. Some improvement is made by assuming that only $\rho = \sigma_B^2/\left(\sigma_e^2+\sigma_B^2\right)$ is known while σ^2 is unknown. In particular, it follows from Section 11.1 that the model with both fixed block effects δ and random block effects β, i.e.,

$$Y = J\mu + X\delta + Z\tau + \eta,$$

provides an estimate of $\sigma^2(1-\rho) = \sigma_e^2$. This estimate is $\hat{\sigma}_e^2$, the mean squared error for the fixed block effect model of Section 9.4.

The key results are that, under model (1),

$$\hat{\sigma}_e^2/\sigma_e^2(1-\rho) \sim \chi^2(rt-b-t+1)$$

and $\hat{\sigma}_e^2$ is independent of $\hat{\tau}$. We show the independence and leave the distributional result to the reader:

Let P be the perpendicular projection operator onto $C(X,Z)$ so

$$\hat{\sigma}_e^2 = \frac{Y'(I-P)Y}{rt-b-t+1}.$$

Independence follows from Theorem 1.2.3 upon observing that

$$\begin{aligned}\mathrm{Cov}\big((I-P)Y,(A-A_0)Y\big) &= \sigma^2(I-P)\left[(1-\rho)I+k\rho M\right](A-A_0)\\ &= \sigma^2(1-\rho)(I-P)(A-A_0)+\sigma^2 k\rho(I-P)M(A-A_0)\\ &= 0.\end{aligned}$$

The last equality holds because $C(A-A_0) \subset C(X,Z) = C(P)$ so that $(I-P)(A-A_0) = 0$ and $(I-P)M = 0$.

A test of model (6) versus model (1) can be based on

$$\frac{r(1-\rho)+\lambda t\rho}{\hat{\sigma}_e^2\left[(1-\rho)+k\rho\right]}\,\frac{\hat{\tau}'(I-P_t)\hat{\tau}}{t-1} \sim F(t-1, rt-t-b-1). \tag{9}$$

This is true because the lefthand side equals

$$\frac{Y'(A-A_0)'V^{-1}(A-A_0)Y/\sigma^2(t-1)}{\hat{\sigma}_e^2/\sigma^2(1-\rho)},$$

which has the appropriate F distribution under H_0. To see the equality of the two statistics, examine the third to the last equality given earlier in the simplification of $Y'(A-A_0)'V^{-1}(A-A_0)Y$. Similarly, tests and confidence intervals can be based on

$$\frac{\xi'\hat{\tau}-\xi'\tau}{\sqrt{\hat{\sigma}_e^2\left[(1-\rho)+k\rho\right]/\left[r(1-\rho)+\lambda t\rho\right]}} \sim t(rt-t-b+1). \tag{10}$$

This uses the second form for $\mathrm{Var}(\xi'\hat{\tau})$ given earlier. To actually use (9) and (10), we need to estimate $\rho = \sigma_B^2/\left(\sigma_e^2+\sigma_B^2\right)$. If we estimate σ_B^2 and take $\hat{\rho} = \hat{\sigma}_B^2/\left(\hat{\sigma}_e^2+\hat{\sigma}_B^2\right)$, the inferential procedures will be identical to those based on (7) and (8), except that they will be based on the more realistic F and t distributions rather than the χ^2 and normal. Thus we have replaced the traditional analysis, which does not account for the estimation of either of the two unknown parameters σ_e^2 and σ_B^2, with an analysis that does not account for the estimation of only one parameter, ρ.

12.11.5 Estimation of Variance Components

The traditional analysis of a BIB with recovery of interblock information uses the variance component estimates of Henderson's Method 3, cf. Section 9. The estimate of σ_e^2 is just that described in the previous subsection. To estimate σ_B^2, let P_τ be the perpendicular projection operator onto $C(Z)$ and recall that P is the perpendicular projection operator onto $C(X,Z)$. Using Henderson's Method 3,

$$\hat{\sigma}_B^2 = \frac{\left[Y'(P-P_\tau)Y - \hat{\sigma}_e^2\,\mathrm{tr}(P-P_\tau)\right]}{\mathrm{tr}\left[X'(P-P_\tau)X\right]}.$$

All of these terms are easily computed. $Y'(P-P_\tau)Y = Y'PY - YP_\tau Y$. $Y'PY$ is available from the fixed block effect analysis. In particular,

$$Y'PY = SS(\text{Grand Mean}) + SS(\text{Blocks}) + SS(\text{Treatments After Blocks})$$

and

$$Y'P_\tau Y = SS(\text{Grand Mean}) + SS(\text{Treatments}),$$

where SS(Treatments) is just the sum of squares from a standard one-way ANOVA that ignores the blocks and the covariance structure. The term $\mathrm{tr}(P-P_\tau)$ is simply $b-1$. It is shown later that $\mathrm{tr}[X'(P-P_\tau)X] = t(r-1)$; thus

$$\hat{\sigma}_B^2 = \frac{SS(\text{Blocks after Treatments}) - \hat{\sigma}_e^2\,(b-1)}{t(r-1)}.$$

To see that $\mathrm{tr}\left[X'(P-P_\tau)X\right] = t(r-1)$, note that $\mathrm{tr}\left[X'(P-P_\tau)X\right] = \mathrm{tr}(X'PX) - \mathrm{tr}(X'P_\tau X) = \mathrm{tr}(X'X) - \mathrm{tr}(X'P_\tau X)$. However, $X'X = kI_b$, so $\mathrm{tr}(X'X) = bk = rt$. The trace of $X'P_\tau X$ is more complicated. From the one-way ANOVA, for any vector Y,

$$P_\tau Y = [t_{ij}], \qquad \text{where } t_{ij} = \frac{1}{r}\sum_{i\in A_j} y_{ij}.$$

The matrix $X = [X_1, \ldots, X_b]$ has

$$X_m = [v_{ij}], \qquad \text{where } v_{ij} = \delta_{im},$$

for $m = 1, \ldots, b$; so applying P_τ to X_m gives

$$P_\tau X_m = [t_{ij}], \qquad \text{where } t_{ij} = \frac{1}{r}\sum_{i\in A_j} \delta_{im} = \frac{1}{r}\delta_m(A_j).$$

Recall that A_j is the set of indices for blocks that include treatment j so that $\delta_m(A_j)$ is 1 if block m contains treatment j, and 0 otherwise. This occurs if and only if treatment j is in block m, so $\delta_m(A_j) = \delta_j(D_m)$. Again, D_m is the set of indices for the treatments contained in block m. It follows that

$$\begin{aligned}
X_m' P_\tau X_m &= [P_\tau X_m]'\,[P_\tau X_m] \\
&= \sum_{j=1}^{t}\sum_{i\in A_j} \frac{1}{r^2}\delta_m(A_j) \\
&= \frac{1}{r^2}\sum_{j=1}^{t}\delta_m(A_j)\sum_{i\in A_j} 1 \\
&= \frac{1}{r}\sum_{j=1}^{t}\delta_m(A_j) \\
&= \frac{1}{r}\sum_{j=1}^{t}\delta_j(D_m) \\
&= \frac{k}{r},
\end{aligned}$$

and therefore

$$\mathrm{tr}\left(X'P_\tau X\right) = \sum_{m=1}^{b} X_m' P_\tau X_m = \frac{bk}{r} = \frac{rt}{r} = t.$$

Combining results gives

$$\operatorname{tr}\left[X'(P-P_\tau)X\right] = rt - t = t(r-1).$$

Exercise 12.10 Do an interblock analysis of the BIB data of Example 9.4.1.

Exercise 12.11 Find the REML estimates of σ_e^2 and σ_B^2.