Chapter 3
Testing

We will consider two approaches to testing linear models. The approaches are identical in that a test under either approach is a well-defined test under the other. The two methods differ only conceptually. One approach is that of testing models; the other approach involves testing linear parametric functions.

Section 1 discusses the notion that a linear model depends fundamentally on $C(X)$ and that the vector of parameters β is of secondary importance. Section 2 discusses testing different models against each other. Section 3 discusses testing linear functions of the β vector. Section 4 presents a brief discussion of the relative merits of testing models versus testing linear parametric functions. Section 5 examines the problem of testing parametric functions that put a constraint on a given subspace of $C(X)$. In particular, Section 5 establishes that estimates and tests can be obtained by using the projection operator onto the subspace. This result is valuable when using one-way ANOVA methods to analyze balanced multifactor ANOVA models. Section 6 considers the problem of breaking sums of squares into independent components. This is a general discussion that relates to breaking ANOVA treatment sums of squares into sums of squares for orthogonal contrasts and also relates to the issue of writing multifactor ANOVA tables with independent sums of squares. Section 7 discusses the construction of confidence regions for estimable linear parametric functions. Section 8 presents testing procedures for the generalized least squares model of Section 2.7.

3.1 More About Models

For estimation in the model

$$Y = X\beta + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 I,$$

we have found that the crucial item needed is M, the perpendicular projection operator onto $C(X)$. For convenience, we will call $C(X)$ the *estimation space* and

R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer Texts in Statistics, DOI 10.1007/978-1-4419-9816-3_3,

$C(X)^\perp$ the *error space*. $I - M$ is the perpendicular projection operator onto the error space. In a profound sense, any two linear models with the same estimation space are the same model. For example, any two such models will give the same predicted values for the observations and the same estimate of σ^2. Suppose we have two linear models for a vector of observations, say $Y = X_1\beta_1 + e_1$ and $Y = X_2\beta_2 + e_2$ with $C(X_1) = C(X_2)$. For these alternative parameterizations, i.e., reparameterizations, M does not depend on which of X_1 or X_2 is used; it depends only on $C(X_1)[= C(X_2)]$. Thus, the MSE does not change, and the least squares estimate of $\mathrm{E}(Y)$ is $\hat{Y} = MY = X_1\hat{\beta}_1 = X_2\hat{\beta}_2$. In fact, we could simply write the original model as

$$\mathrm{E}(Y) \in C(X), \quad \mathrm{Cov}(Y) = \sigma^2 I.$$

The expected values of the observations are fundamental to estimation in linear models. Identifiable parameteric functions are functions of the expected values. Attention is often restricted to estimable functions, i.e., functions $\rho' X\beta$ where $X\beta = \mathrm{E}(Y)$. The key idea in estimability is restricting estimation to linear combinations of the rows of $\mathrm{E}(Y)$. $\mathrm{E}(Y)$ depends only on the choice of $C(X)$, whereas the vector β depends on the particular choice of X. Consider again the two models discussed above. If $\lambda_1'\beta_1$ is estimable, then $\lambda_1'\beta_1 = \rho' X_1\beta_1 = \rho' E(Y)$ for some ρ. This estimable function is the same linear combination of the rows of $\mathrm{E}(Y)$ as $\rho' E(Y) = \rho' X_2\beta_2 = \lambda_2'\beta_2$. These are really the same estimable function, but they are written with different parameters. This estimable function has a unique least squares estimate, $\rho' MY$.

EXAMPLE 3.1.1. *One-Way ANOVA.*
Two parameterizations for a one-way ANOVA are commonly used. They are

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

and

$$y_{ij} = \mu_i + e_{ij}.$$

It is easily seen that these models determine the same estimation space. The estimates of σ^2 and $\mathrm{E}(y_{ij})$ are identical in the two models. One convenient aspect of these models is that the relationships between the two sets of parameters are easily identified. In particular,

$$\mu_i = \mathrm{E}(y_{ij}) = \mu + \alpha_i.$$

It follows that the mean of the μ_is equals μ plus the mean of the α_is, i.e., $\bar{\mu}_\cdot = \mu + \bar{\alpha}_\cdot$. It also follows that $\mu_1 - \mu_2 = \alpha_1 - \alpha_2$, etc. The parameters in the two models are different, but they are related. Any estimable function in one model determines a corresponding estimable function in the other model. These functions have the same estimate. Chapter 4 contains a detailed examination of these models.

EXAMPLE 3.1.2. *Simple Linear Regression.*
The models

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

and

$$y_i = \gamma_0 + \gamma_1 (x_i - \bar{x}_\cdot) + e_i$$

have the same estimation space ($\bar{x}_\cdot$ is the mean of the x_is). Since

$$\beta_0 + \beta_1 x_i = \mathrm{E}(y_i) = \gamma_0 + \gamma_1 (x_i - \bar{x}_\cdot) \tag{1}$$

for all i, it is easily seen from averaging over i that

$$\beta_0 + \beta_1 \bar{x}_\cdot = \gamma_0 .$$

Substituting $\beta_0 = \gamma_0 - \beta_1 \bar{x}_\cdot$ into (1) leads to

$$\beta_1 (x_i - \bar{x}_\cdot) = \gamma_1 (x_i - \bar{x}_\cdot)$$

and, if the x_is are not all identical,

$$\beta_1 = \gamma_1 .$$

These models are examined in detail in Section 6.1.

When the estimation spaces $C(X_1)$ and $C(X_2)$ are the same, write $X_1 = X_2 T$ to get

$$X_1 \beta_1 = X_2 T \beta_1 = X_2 \beta_2 . \tag{2}$$

Estimable functions are equivalent in the two models: $\Lambda_1' \beta_1 = P' X_1 \beta_1 = P' X_2 \beta_2 = \Lambda_2' \beta_2$. It also follows from equation (2) that the parameterizations must satisfy the relation

$$\beta_2 = T \beta_1 + v \tag{3}$$

for some $v \in C(X_2')^\perp$. In general, neither of the parameter vectors β_1 or β_2 is uniquely defined but, to the extent that either parameter vector is defined, equation (3) establishes the relationship between them. A unique parameterization for, say, the X_2 model occurs if and only if $X_2' X_2$ is nonsingular. In such a case, the columns of X_2 form a basis for $C(X_2)$, so the matrix T is uniquely defined. In this case, the vector v must be zero because $C(X_2')^\perp = \{0\}$. An alternative and detailed presentation of equivalent linear models, both the reparameterizations considered here and the equivalences between constrained and unconstrained models considered in subsequent sections, is given by Peixoto (1993).

Basically, the β parameters in

$$Y = X\beta + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 I$$

are either a convenience or a nuisance, depending on what we are trying to do. Having $\mathrm{E}(e) = 0$ gives $\mathrm{E}(Y) = X\beta$, but since β is unknown, this is merely saying that $\mathrm{E}(Y)$ is *some* linear combination of the columns of X. The essence of the model is that

$$\mathrm{E}(Y) \in C(X), \quad \mathrm{Cov}(Y) = \sigma^2 I.$$

As long as we do not change $C(X)$, we can change X itself to suit our convenience.

3.2 Testing Models

In this section, the basic theory for testing a linear model against a reduced model is presented. A generalization of the basic procedure is also presented.

Testing in linear models typically reduces to putting a constraint on the estimation space. We start with a (full) model that we know (assume) to be valid,

$$Y = X\beta + e, \quad e \sim N(0, \sigma^2 I). \tag{1}$$

Our wish is to reduce this model, i.e., we wish to know if some simpler model gives an acceptable fit to the data. Consider whether the model

$$Y = X_0\gamma + e, \quad e \sim N(0, \sigma^2 I), \quad C(X_0) \subset C(X) \tag{2}$$

is acceptable. Clearly, if model (2) is correct, then model (1) is also correct. The question is whether (2) is correct.

The procedure of testing full and reduced models is a commonly used method in statistics.

EXAMPLE 3.2.0.

(a) *One-Way ANOVA.*
The full model is

$$y_{ij} = \mu + \alpha_i + e_{ij}.$$

To test for no treatment effects, i.e., to test that the α_is are extraneous, the reduced model simply eliminates the treatment effects. The reduced model is

$$y_{ij} = \gamma + e_{ij}.$$

Additionally, consider testing $H_0 : \alpha_1 - \alpha_3 = 0$ in Example 1.0.2. The full model is

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{31} \\ e_{32} \end{bmatrix}.$$

We can rewrite this as

$$Y = \mu \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + e.$$

If we impose the constraint $H_0 : \alpha_1 - \alpha_3 = 0$, i.e., $\alpha_1 = \alpha_3$, we get

$$Y = \mu J + \alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_1 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + e,$$

or

$$Y = \mu J + \alpha_1 \left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right) + \alpha_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + e,$$

or

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{31} \\ e_{32} \end{bmatrix}.$$

This is the reduced model determined by $H_0 : \alpha_1 - \alpha_3 = 0$. However, the parameters μ, α_1, and α_2 no longer mean what they did in the full model.

(b) *Multiple Regression.*
Consider the full model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i.$$

For a simultaneous test of whether the variables x_1 and x_3 are adding significantly to the explanatory capability of the regression model, simply eliminate the variables x_1 and x_3 from the model. The reduced model is

$$y_i = \gamma_0 + \gamma_2 x_{i2} + e_i.$$

Now write the original model matrix as $X = [J, X_1, X_2, X_3]$, so

$$Y = [J, X_1, X_2, X_3] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + e = \beta_0 J + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e.$$

Consider the hypothesis $H_0 : \beta_2 - \beta_3 = 0$ or $H_0 : \beta_2 = \beta_3$. The reduced model is

$$\begin{aligned} Y &= \beta_0 J + \beta_1 X_1 + \beta_2 X_2 + \beta_2 X_3 + e \\ &= \beta_0 J + \beta_1 X_1 + \beta_2 (X_2 + X_3) + e \\ &= [J, X_1, X_2 + X_3] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + e. \end{aligned}$$

However, these β parameters no longer mean what they did in the original model, so it is better to write the model as

$$Y = [J, X_1, X_2 + X_3] \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} + e.$$

The distribution theory for testing models is given in Theorem 3.2.1. Before stating those results, we discuss the intuitive background of the test based only on the assumptions about the first two moments. Let M and M_0 be the perpendicular projection operators onto $C(X)$ and $C(X_0)$, respectively. Note that with $C(X_0) \subset C(X)$, $M - M_0$ is the perpendicular projection operator onto the orthogonal complement of $C(X_0)$ with respect to $C(X)$, that is, onto $C(M - M_0) = C(X_0)^{\perp}_{C(X)}$, see Theorems B.47 and B.48.

Under model (1), the estimate of $\mathrm{E}(Y)$ is MY. Under model (2), the estimate is $M_0 Y$. Recall that the validity of model (2) implies the validity of model (1); so if model (2) is true, MY and $M_0 Y$ are estimates of the same quantity. This suggests that the difference between the two estimates, $MY - M_0 Y = (M - M_0)Y$, should be reasonably small. Under model (2), the difference is just error because $\mathrm{E}[(M - M_0)Y] = 0$.

On the other hand, a large difference between the estimates suggests that MY and $M_0 Y$ are estimating different things. By assumption, MY is always an estimate of $\mathrm{E}(Y)$; so $M_0 Y$ must be estimating something different, namely, $M_0 \mathrm{E}(Y) \neq \mathrm{E}(Y)$. If $M_0 Y$ is not estimating $\mathrm{E}(Y)$, model (2) cannot be true because model (2) implies that $M_0 Y$ is an estimate of $\mathrm{E}(Y)$.

The decision about whether model (2) is appropriate hinges on deciding whether the vector $(M - M_0)Y$ is large. An obvious measure of the size of $(M - M_0)Y$ is its squared length, $[(M - M_0)Y]'[(M - M_0)Y] = Y'(M - M_0)Y$. However, the length of $(M - M_0)Y$ is also related to the relative sizes of $C(M)$ and $C(M_0)$. It is convenient (not crucial) to adjust for this factor. As a measure of the size of $(M - M_0)Y$, we use the value

$$Y'(M - M_0)Y \big/ r(M - M_0).$$

Even though we have an appropriate measure of the size of $(M-M_0)Y$, we still need some idea of how large the measure will be both when model (2) is true and when model (2) is not true. Using only the assumption that model (1) is true, Theorem 1.3.2 implies that

$$\mathrm{E}[Y'(M-M_0)Y/r(M-M_0)] = \sigma^2 + \beta'X'(M-M_0)X\beta/r(M-M_0).$$

If model (2) is also true, $X\beta = X_0\gamma$ and $(M-M_0)X_0 = 0$; so the expected value of $Y'(M-M_0)Y/r(M-M_0)$ is σ^2. If σ^2 were known, our intuitive argument would be complete. If $Y'(M-M_0)Y/r(M-M_0)$ is not much larger than σ^2, then we have observed something that is consistent with the validity of model (2). Values that are much larger than σ^2 indicate that model (2) is false because they suggest that $\beta'X'(M-M_0)X\beta/r(M-M_0) > 0$.

Typically, we do not know σ^2, so the obvious thing to do is to estimate it. Since model (1) is assumed to be true, the obvious estimate is the $MSE = Y'(I-M)Y/r(I-M)$. Now, values of $Y'(M-M_0)Y/r(M-M_0)$ that are much larger than MSE cause us to doubt the validity of model (2). Equivalently, values of the *test statistic*

$$\frac{Y'(M-M_0)Y/r(M-M_0)}{MSE}$$

that are considerably larger than 1 cause precisely the same doubts.

We now examine the behavior of this test statistic when model (2) is not correct but model (1) is. $Y'(M-M_0)Y/r(M-M_0)$ and MSE are each estimates of their expected values, so the test statistic obviously provides an estimate of the ratio of their expected values. Recalling $\mathrm{E}[Y'(M-M_0)Y/r(M-M_0)]$ from above and that $\mathrm{E}(MSE) = \sigma^2$, the test statistic gives an estimate of $1 + \beta'X'(M-M_0)X\beta/r(M-M_0)\sigma^2$. The term $\beta'X'(M-M_0)X\beta$ is crucial to evaluating the behavior of the test statistic when model (1) is valid but model (2) is not, cf. the noncentrality parameter in Theorem 3.2.1, part *i*. Note that $\beta'X'(M-M_0)X\beta = [X\beta - M_0X\beta]'[X\beta - M_0X\beta]$ is the squared length of the difference between $X\beta$ (i.e., E(Y)) and $M_0X\beta$ (the projection of $X\beta$ onto $C(X_0)$). If $X\beta - M_0X\beta$ is large (relative to σ^2), then model (2) is very far from being correct, and the test statistic will tend to be large. On the other hand, if $X\beta - M_0X\beta$ is small (relative to σ^2), then model (2), although not correct, is a good approximation to the correct model. In this case the test statistic will tend to be a little larger than it would be if model (2) were correct, but the effect will be very slight. In other words, if $\beta'X'(M-M_0)X\beta/r(M-M_0)\sigma^2$ is small, it is unreasonable to expect any test to work very well.

One can think about the geometry of all this in three dimensions. As in Section 2.2, consider a rectangular table. Take one corner of the table to be the origin. Take $C(X)$ as the two-dimensional subspace determined by the surface of the table and take $C(X_0)$ to be a one-dimensional subspace determined by an edge of the table that includes the origin. Y can be any vector originating at the origin, i.e., any point in three-dimensional space. The full model (1) says that $\mathrm{E}(Y) = X\beta$, which just says that $\mathrm{E}(Y)$ is somewhere on the surface of the table. The reduced model (2) says that $\mathrm{E}(Y)$ is somewhere on the $C(X_0)$ edge of the table. $MY = X\hat{\beta}$ is the perpendicular projection of Y onto the table surface. $M_0Y = X_0\hat{\gamma}$ is the perpendicular projection of Y onto the $C(X_0)$ edge of the table. The residual vector $(I-M)Y$ is the perpendicular projection of Y onto the vertical line through the origin.

If MY is close to the $C(X_0)$ edge of the table, it must be close to M_0Y. This is the behavior we would expect if the reduced model is true, i.e., if $X\beta = X_0\gamma$. The difference between the two estimates, $MY - M_0Y$, is a vector that is on the table, but perpendicular to the $C(X_0)$ edge. In fact, the table edge through the origin that is perpendicular to the $C(X_0)$ edge is the orthogonal complement of $C(X_0)$ with respect to $C(X)$, that is, it is $C(X_0)^{\perp}_{C(X)} = C(M - M_0)$. The difference between the two estimates is $MY - M_0Y = (M - M_0)Y$, which is the perpendicular projection of Y onto $C(M - M_0)$. If $(M - M_0)Y$ is large, it suggests that the reduced model is not true. To decide if $(M - M_0)Y$ is large, we find its average (mean) squared length, where the average is computed relative to the dimension of $C(M - M_0)$, and compare that to the averaged squared length of the residual vector $(I-M)Y$ (i.e., the MSE). In our three-dimensional example, the dimensions of both $C(M - M_0)$ and $C(I-M)$ are 1. If $(M - M_0)Y$ is, on average, much larger than $(I-M)Y$, we reject the reduced model.

If the true (but unknown) $X\beta$ happens to be far from the $C(X_0)$ edge of the table, it will be very easy to see that the reduced model is not true. This occurs because MY will be near $X\beta$, which is far from anything in $C(X_0)$ so, in particular, it will be far from M_0Y. Remember that the meaning of "far" depends on σ^2 which is estimated by the MSE. On the other hand, if $X\beta$ happens to be near, but not on, the $C(X_0)$ edge of the table, it will be very hard to tell that the reduced model is not true because MY and M_0Y will tend to be close together. On the other hand, if $X\beta$ is near, but not on, the $C(X_0)$ edge of the table, using the incorrect reduced model may not create great problems.

To this point, the discussion has been based entirely on the assumptions $Y = X\beta + e$, $\mathrm{E}(e) = 0$, $\mathrm{Cov}(e) = \sigma^2 I$. We now quantify the precise behavior of the test statistic for normal errors.

Theorem 3.2.1. Consider a full model

$$Y = X\beta + e, \quad e \sim N(0, \sigma^2 I)$$

that holds for some values of β and σ^2 and a reduced model

$$Y = X_0\gamma + e, \quad e \sim N(0, \sigma^2 I), \quad C(X_0) \subset C(X).$$

(i) If the full model is true,

$$\frac{Y'(M - M_0)Y/r(M - M_0)}{Y'(I-M)Y/r(I-M)} \sim F\left(r(M - M_0), r(I-M), \beta'X'(M - M_0)X\beta/2\sigma^2\right).$$

(ii) If the reduced model is true,

$$\frac{Y'(M-M_0)Y/r(M-M_0)}{Y'(I-M)Y/r(I-M)} \sim F(r(M-M_0), r(I-M), 0).$$

When the full model is true, this distribution holds only if the reduced model is true.

PROOF.

(i) Since M and M_0 are the perpendicular projection matrices onto $C(X)$ and $C(X_0)$, $M-M_0$ is the perpendicular projection matrix onto $C(M-M_0)$, cf. Theorem B.47. As in Section 2.6,

$$\frac{Y'(I-M)Y}{\sigma^2} \sim \chi^2(r(I-M))$$

and from Theorem 1.3.3 on the distribution of quadratic forms,

$$\frac{Y'(M-M_0)Y}{\sigma^2} \sim \chi^2\left(r(M-M_0), \beta'X'(M-M_0)X\beta/2\sigma^2\right).$$

Theorem 1.3.7 establishes that $Y'(M-M_0)Y$ and $Y'(I-M)Y$ are independent because

$$\begin{aligned}(M-M_0)(I-M) &= M-M_0-M+M_0M \\ &= M-M_0-M+M_0 = 0.\end{aligned}$$

Finally, part (i) of the theorem holds by Definition C.3.

(ii) It suffices to show that $\beta'X'(M-M_0)X\beta = 0$ if and only if $\mathrm{E}(Y) = X_0\gamma$ for some γ.

$\Leftarrow$ If $\mathrm{E}(Y) = X_0\gamma$, we have $\mathrm{E}(Y) = X\beta$ for some β because $C(X_0) \subset C(X)$. In particular, $\beta'X'(M-M_0)X\beta = \gamma'X_0'(M-M_0)X_0\gamma$, but since $(M-M_0)X_0 = X_0 - X_0 = 0$, we have $\beta'X'(M-M_0)X\beta = 0$.

$\Rightarrow$ If $\beta'X'(M-M_0)X\beta = 0$, then $[(M-M_0)X\beta]'[(M-M_0)X\beta] = 0$. Since for any x, $x'x = 0$ if and only if $x = 0$, we have $(M-M_0)X\beta = 0$ or $X\beta = M_0X\beta = X_0(X_0'X_0)^-X_0'X\beta$. Taking $\gamma = (X_0'X_0)^-X_0'X\beta$, we have $\mathrm{E}(Y) = X_0\gamma$. □

People typically reject the hypothesis

$$H_0 : \mathrm{E}(Y) = X_0\gamma \qquad \text{for some } \gamma,$$

for large observed values of the test statistic. The informal second moment arguments given prior to Theorem 3.2.1 suggest rejecting large values and the existence a positive noncentrality parameter in Theorem 3.2.1(i) would shift the (central) F distribution to the right which also suggests rejecting large values. Both of these arguments depend on the full model being true. Theorem 3.2.1(ii) provides a distribution for the test statistic under the reduced (null) model, so under the conditions of the theorem this test of

$$H_0 : \mathrm{E}(Y) \in C(X_0)$$

rejects H_0 at level α if

$$\frac{Y'(M-M_0)Y/r(M-M_0)}{Y'(I-M)Y/r(I-M)} > F(1-\alpha, r(M-M_0), r(I-M)).$$

P values are then reported as the probability from an $F(r(M-M_0), r(I-M))$ distribution of being at least as large as the observed value of the test statistic.

This test procedure is "non-Fisherian" in that it assumes more than just the null (reduced) model being true. The decision on when to reject the null model depends on the full model being true. In fact, test statistic values near 0 (reported P values near 1) can be just as interesting as large test statistic values (reported P values near 0), although large reported P values often need to be closer to 1 to be interesting than small reported P values need to be close to 0. Personally, I don't consider these reported P values to be real P values, although they are not without their uses. *Appendix F discusses the significance of small test statistics and some foundational issues related to this test.* For those willing to assume that the full model is true, this test is the *uniformly most powerful invariant (UMPI) test* and the *generalized likelihood ratio test* (see Lehmann, 1986, Chapter 7 and Exercise 3.1).

In practice it is often easiest to use the following approach to obtain the test statistic: Observe that $(M-M_0) = (I-M_0) - (I-M)$. If we can find the error sums of squares, $Y'(I-M_0)Y$ from the model $Y = X_0\gamma + e$ and $Y'(I-M)Y$ from the model $Y = X\beta + e$, then the difference is $Y'(I-M_0)Y - Y'(I-M)Y = Y'(M-M_0)Y$, which is the numerator sum of squares for the F test. Unless there is some simplifying structure to the model matrix (as in cases we will examine later), it is usually easier to obtain the error sums of squares for the two models than to find $Y'(M-M_0)Y$ directly.

EXAMPLE 3.2.2. Consider the model matrix given at the end of this example. It is for the unbalanced analysis of variance $y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}$, where $i = 1,2,3$, $j = 1,2,3$, $k = 1,\dots,N_{ij}$, $N_{11} = N_{12} = N_{21} = 3$, $N_{13} = N_{22} = N_{23} = N_{31} = N_{32} = N_{33} = 2$. Here we have written $Y = X\beta + e$ with $Y = [y_{111}, y_{112}, \dots, y_{332}]'$, $\beta = [\mu, \alpha_1, \alpha_2, \alpha_3, \eta_1, \eta_2, \eta_3]'$, and $e = [e_{111}, e_{112}, \dots, e_{332}]'$. We can now test to see if the model $y_{ijk} = \mu + \alpha_i + e_{ijk}$ is an adequate representation of the data simply by dropping the last three columns from the model matrix. We can also test $y_{ijk} = \mu + e_{ijk}$ by dropping the last six columns of the model matrix. In either case, the test is based on comparing the error sum of squares for the reduced model with that of the full model.

$$X = \begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}.$$

3.2.1 A Generalized Test Procedure

Before considering tests of parametric functions, we consider a generalization of the test procedure outlined earlier. Assume that the model $Y = X\beta + e$ is correct. We want to test the adequacy of a model $Y = X_0\gamma + Xb + e$, where $C(X_0) \subset C(X)$ and Xb is some known vector. In generalized linear model terminology, Xb is called an *offset*.

EXAMPLE 3.2.3. *Multiple Regression.* Consider the model

$$Y = [J, X_1, X_2, X_3] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + e = \beta_0 J + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e.$$

To test $H_0 : \beta_2 = \beta_3 + 5, \beta_1 = 0$, write the reduced model as

$$\begin{aligned} Y &= \beta_0 J + (\beta_3 + 5)X_2 + \beta_3 X_3 + e \\ &= \beta_0 J + \beta_3(X_2 + X_3) + 5X_2 + e \qquad (3) \\ &= [J, X_2 + X_3] \begin{bmatrix} \beta_0 \\ \beta_3 \end{bmatrix} + 5X_2 + e. \end{aligned}$$

Alternatively, we could write the reduced model as

$$\begin{aligned} Y &= \beta_0 J + \beta_2 X_2 + (\beta_2 - 5)X_3 + e \\ &= \beta_0 J + \beta_2 (X_2 + X_3) - 5X_3 + e \\ &= [J, X_2 + X_3] \begin{bmatrix} \beta_0 \\ \beta_2 \end{bmatrix} - 5X_3 + e. \end{aligned} \tag{4}$$

We will see that both reduced models lead to the same test.

The model $Y = X\beta + e$ can be rewritten as $Y - Xb = X\beta - Xb + e$. Since $Xb \in C(X)$, this amounts to a reparameterization, $Y - Xb = X\beta^* + e$, where $\beta^* = \beta - b$. Since Xb is known, $Y - Xb$ is still an observable random variable.

The reduced model $Y = X_0\gamma + Xb + e$ can be rewritten as $Y - Xb = X_0\gamma + e$. The question of testing the adequacy of the reduced model is now a straightforward application of our previous theory. The distribution of the test statistic is

$$\frac{(Y - Xb)'(M - M_0)(Y - Xb)/r(M - M_0)}{(Y - Xb)'(I - M)(Y - Xb)/r(I - M)} \sim F\left(r(M - M_0), r(I - M), \beta^{*\prime} X'(M - M_0)X\beta^*/2\sigma^2\right).$$

The noncentrality parameter is zero if and only if $0 = \beta^{*\prime}X'(M - M_0)X\beta^* = [(M - M_0)(X\beta - Xb)]'[(M - M_0)(X\beta - Xb)]$, which occurs if and only if $(M - M_0)(X\beta - Xb) = 0$ or $X\beta = M_0(X\beta - Xb) + Xb$. The last condition is nothing more or less than that the reduced model is valid with $\gamma = (X_0'X_0)^- X_0(X\beta - Xb)$, a fixed unknown parameter.

Note also that, since $(I - M)X = 0$, in the denominator of the test statistic $(Y - Xb)'(I - M)(Y - Xb) = Y'(I - M)Y$, which is the SSE from the original full model. Moreover, the numerator sum of squares is

$$(Y - Xb)'(M - M_0)(Y - Xb) = (Y - Xb)'(I - M_0)(Y - Xb) - Y'(I - M)Y,$$

which can be obtained by subtracting the SSE of the original full model from the SSE of the reduced model. To see this, write $I - M_0 = (I - M) + (M - M_0)$.

EXAMPLE 3.2.3 CONTINUED. The numerator sum of squares for testing model (4) is $(Y + 5X_3)'(M - M_0)(Y + 5X_3)$. But $(M - M_0)[X_2 + X_3] = 0$, so, upon observing that $5X_3 = -5X_2 + 5[X_2 + X_3]$,

$$\begin{aligned} &(Y + 5X_3)'(M - M_0)(Y + 5X_3) \\ &= (Y - 5X_2 + 5[X_2 + X_3])'(M - M_0)(Y - 5X_2 + 5[X_2 + X_3]) \\ &= (Y - 5X_2)'(M - M_0)(Y - 5X_2), \end{aligned}$$

which is the numerator sum of squares for testing model (3). In fact, models (3) and (4) are equivalent because the only thing different about them is that one uses $5X_2$ and the other uses $-5X_3$; but the only difference between these terms is $5[X_2 + X_3] \in C(X_0)$.

The phenomenon illustrated in this example is a special case of a general result. Consider the model $Y = X_0\gamma + Xb + e$ for some unknown γ and known b and suppose $X(b - b_*) \in C(X_0)$ for known b_*. The model $\mathrm{E}(Y) = X_0\gamma + Xb$ holds if and only if $\mathrm{E}(Y) = X_0\gamma + X(b - b_*) + Xb_*$, which holds if and only if $\mathrm{E}(Y) = X_0\gamma_* + Xb_*$ for some unknown γ_*.

Exercise 3.1 (a) Show that the F test developed in the first part of this section is equivalent to the (generalized) likelihood ratio test for the reduced versus full models, cf. Casella and Berger (2002, Subsection 8.2.1). (b) Find an F test for $H_0 : X\beta = X\beta_0$ where β_0 is known. (c) Construct a full versus reduced model test when σ^2 has a known value σ_0^2.

Exercise 3.2 Redo the tests in Exercise 2.2 using the theory of Section 3.2. Write down the models and explain the procedure.

Exercise 3.3 Redo the tests in Exercise 2.3 using the procedures of Section 3.2. Write down the models and explain the procedure.

Hints: (a) Let A be a matrix of zeros, the generalized inverse of A, A^-, can be anything at all because $AA^-A = A$ for any choice of A^-. (b) There is no reason why X_0 cannot be a matrix of zeros.

3.3 Testing Linear Parametric Functions

In this section, the theory of testing linear parametric functions is presented. A basic test procedure and a generalized test procedure are given. These procedures are analogous to the model testing procedures of Section 2. In the course of this presentation, the important concept of the constraint imposed by an hypothesis is introduced. Finally, a class of hypotheses that is rarely used for linear models but commonly used with log-linear models is given along with results that define the appropriate testing procedure.

Consider a general linear model

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

with X an $n \times p$ matrix. A key aspect of this model is that β is allowed to be any vector in $\mathbf{R}^p$. Additionally, consider an hypothesis concerning a linear function, say $\Lambda'\beta = 0$. The null model to be tested is

$$H_0 : Y = X\beta + e \quad \text{and} \quad \Lambda'\beta = 0.$$

We need to find a reduced model that corresponds to this.

The constraint $\Lambda'\beta = 0$ can be stated in an infinite number of ways. Observe that $\Lambda'\beta = 0$ holds if and only if $\beta \perp C(\Lambda)$; so if Γ is another matrix with $C(\Gamma) = C(\Lambda)$, the constraint can also be written as $\beta \perp C(\Gamma)$ or $\Gamma'\beta = 0$.

To identify the reduced model, pick a matrix U such that

$$C(U) = C(\Lambda)^{\perp},$$

then $\Lambda'\beta = 0$ if and only if $\beta \perp C(\Lambda)$ if and only if $\beta \in C(U)$, which occurs if and only if for some vector γ,

$$\beta = U\gamma. \tag{2}$$

Substituting (2) into the linear model gives the reduced model

$$Y = XU\gamma + e$$

or, letting $X_0 \equiv XU$,

$$Y = X_0\gamma + e. \tag{3}$$

Note that $C(X_0) \subset C(X)$. If $e \sim N(0, \sigma^2 I)$, the reduced model (3) allows us to test $\Lambda'\beta = 0$ by applying the results of Section 2. If $C(X_0) = C(X)$, the constraint involves only a reparameterization and there is nothing to test. In other words, if $C(X_0) = C(X)$, then $\Lambda'\beta = 0$ involves only arbitrary side conditions that do not affect the model. Moreover, the reduced model does not depend on $\mathrm{Cov}(Y)$ or the exact distribution of e, it only depends on $\mathrm{E}(Y) = X\beta$ and the constraint $\Lambda'\beta = 0$.

EXAMPLE 3.3.1. Consider the one-way analysis of variance model

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \beta + e,$$

where $\beta = (\mu, \ \alpha_1, \ \alpha_2, \ \alpha_3)'$. The parameters in this model are not uniquely defined because the rank of X is less than the number of columns.

Let $\lambda_1' = (0, 1, 0, -1)$. The contrast $\lambda_1'\beta$ is estimable, so the hypothesis $\alpha_1 - \alpha_3 = \lambda_1'\beta = 0$ determines an estimable constraint. To obtain $C(\lambda_1)^{\perp} = C(U)$, one can pick

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

which yields

$$XU = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}. \tag{4}$$

This is a real restriction on $C(X)$, i.e., $C(XU) \neq C(X)$.

Let $\lambda_2' = (0,1,1,1)$. A nonestimable linear constraint for a one-way analysis of variance is that

$$\alpha_1 + \alpha_2 + \alpha_3 = \lambda_2'\beta = 0.$$

Consider two choices for U with $C(\lambda_2)^\perp = C(U)$, i.e.,

$$U_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & -2 \end{bmatrix} \quad \text{and} \quad U_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -2 \\ 0 & -1 & 1 \end{bmatrix}.$$

These yield

$$XU_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 0 & -2 \end{bmatrix} \quad \text{and} \quad XU_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}.$$

Note that $C(X) = C(XU_1) = C(XU_2)$. The models determined by XU_1 and XU_2 are equivalent linear models but have different parameterizations, say $Y = XU_1\xi_1 + e$ and $Y = XU_2\xi_2 + e$, with $\xi_i = (\xi_{i1}, \xi_{i2}, \xi_{i3})'$. Transforming to the original β parameterization using (2), for example,

$$\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \beta = U_1\xi_1 = \begin{bmatrix} \xi_{11} \\ \xi_{12} + \xi_{13} \\ -\xi_{12} + \xi_{13} \\ -2\xi_{13} \end{bmatrix}.$$

Both ξ_i parameterizations lead to $\alpha_1 + \alpha_2 + \alpha_3 = 0$. Thus both determine the same specific choice for the parameterization of the original one-way analysis of variance model.

Similar results hold for alternative nonidentifiable constraints such as $\alpha_1 = 0$. As will be established later, any nonidentifiable constraint leaves the estimation space unchanged and therefore yields the same estimates of estimable functions.

Now consider the joint constraint $\Lambda_1'\beta = 0$, where

$$\Lambda_1' = \begin{bmatrix} \lambda_1' \\ \lambda_2' \end{bmatrix}.$$

$\lambda_1'\beta$ is a contrast, so it is estimable; therefore $\Lambda_1'\beta$ has estimable aspects. One choice of U with $C(\Lambda_1)^\perp = C(U)$ is

$$U_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -2 \\ 0 & 1 \end{bmatrix}.$$

This gives

$$XU_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix},$$

and the estimation space is the same as in (4), where only the contrast $\lambda_1'\beta$ was assumed equal to zero.

A constraint equivalent to $\Lambda_1'\beta = 0$ is $\Lambda_2'\beta = 0$, where

$$\Lambda_2' = \begin{bmatrix} \lambda_3' \\ \lambda_2' \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

The constraints are equivalent because $C(\Lambda_1) = C(\Lambda_2)$. (Note that $\lambda_3 - \lambda_2 = \lambda_1$.) Neither $\lambda_3'\beta$ nor $\lambda_2'\beta$ is estimable, so separately they would affect only the parameterization of the model. However, $\Lambda_1'\beta = 0$ involves an estimable constraint, so $\Lambda_2'\beta = 0$ also has an estimable part to the constraint. The concept of the estimable part of a constraint will be examined in detail later.

Estimable Constraints

We have established a perfectly general method for identifying the reduced model determined by an arbitrary linear constraint $\Lambda'\beta = 0$, and thus we have a general method for testing $\Lambda'\beta = 0$ by applying the results of Section 2. Next, we will examine the form of the test statistic when $\Lambda'\beta$ is estimable. Afterwards, we present results showing that there is little reason to consider nonestimable linear constraints.

When $\Lambda'\beta$ is estimable, so that $\Lambda' = P'X$ for some P, rather than finding U we can find the numerator projection operator for testing $\Lambda'\beta = 0$ in terms of P and M. Better yet, we can find the numerator sum of squares in terms of Λ and any least squares estimate $\hat{\beta}$. Recall from Section 2 that the numerator sum of squares is $Y'(M - M_0)Y$, where $M - M_0$ is the perpendicular projection operator onto the orthogonal complement of $C(X_0)$ with respect to $C(X)$. In other words, $M - M_0$ is a perpendicular projection operator with $C(M - M_0) = C(X_0)^\perp_{C(X)}$. For testing an estimable parametric hypothesis with $\Lambda' = P'X$, we now show that the perpendicular projection operator onto $C(MP)$ is also the perpendicular projection operator onto the orthogonal complement of $C(X_0)$ with respect to $C(X)$, i.e., that $C(MP) = C(X_0)^{\perp}_{C(X)}$. It follows immediately that the numerator sum of squares in the test is $Y'M_{MP}Y$, where $M_{MP} \equiv MP(P'MP)^{-}P'M$ is the perpendicular projection operator onto $C(MP)$. In particular, from Section 2

$$\frac{Y'M_{MP}Y/r(M_{MP})}{Y'(I-M)Y/r(I-M)} \sim F\left(r(M_{MP}), r(I-M), \beta'X'M_{MP}X\beta/2\sigma^2\right). \qquad (5)$$

Proposition 3.3.2 provides a formal proof of the necessary result. However, after the proof, we give an alternative justification based on finding the reduced model associated with the constraint. This reduced model argument differs from the one given at the beginning of the section in that it only applies to estimable constraints.

Proposition 3.3.2. With U and P defined for $\Lambda'\beta = 0$,

$$C(M - M_0) = C(X_0)^{\perp}_{C(X)} \equiv C(XU)^{\perp}_{C(X)} = C(MP).$$

PROOF. From Section 2, we already know that $C(M - M_0) = C(X_0)^{\perp}_{C(X)}$. Since $X_0 \equiv XU$, we need only establish that $C(XU)^{\perp}_{C(X)} = C(MP)$.

If $x \in C(XU)^{\perp}_{C(X)}$, then $0 = x'XU$, so $X'x \perp C(U)$ and $X'x \in C(\Lambda) = C(X'P)$. It follows that

$$x = Mx = [X(X'X)^{-}]X'x \in C([X(X'X)^{-}]X'P) = C(MP).$$

Conversely, if $x \in C(MP)$, then $x = MPb$ for some b and

$$x'XU = b'P'MXU = b'P'XU = b'\Lambda'U = 0,$$

so $x \in C(XU)^{\perp}_{C(X)}$. $\square$

Exercise 3.4 Show that $\beta'X'M_{MP}X\beta = 0$ if and only if $\Lambda'\beta = 0$.

Earlier, we found the reduced model matrix $X_0 = XU$ directly and then, for $\Lambda' = P'X$, we showed that $C(MP) = C(X_0)^{\perp}_{C(X)}$, which led to the numerator sum of squares. An alternative derivation of the test arrives at $C(M - M_0) = C(X_0)^{\perp}_{C(X)} = C(MP)$ more directly for estimable constraints. The reduced model is

$$Y = X\beta + e \quad \text{and} \quad P'X\beta = 0,$$

or

$$Y = X\beta + e \quad \text{and} \quad P'MX\beta = 0,$$

or

$$\mathrm{E}(Y) \in C(X) \quad \text{and} \quad \mathrm{E}(Y) \perp C(MP),$$

or

$$\mathrm{E}(Y) \in C(X) \cap C(MP)^{\perp}.$$

The reduced model matrix X_0 must satisfy $C(X_0) = C(X) \cap C(MP)^{\perp} \equiv C(MP)^{\perp}_{C(X)}$. It follows immediately that $C(X_0)^{\perp}_{C(X)} = C(MP)$. Moreover, it is easily seen that X_0 can be taken as $X_0 = (I - M_{MP})X$.

Theorem 3.3.3. $\quad C[(I - M_{MP})X] = C(X) \cap C(MP)^{\perp}$.

PROOF. First, assume $x \in C(X)$ and $x \perp C(MP)$. Write $x = Xb$ for some b and note that $M_{MP}x = 0$. It follows that $x = (I - M_{MP})x = (I - M_{MP})Xb$, so $x \in C[(I - M_{MP})X]$.

Conversely, if $x = (I - M_{MP})Xb$ for some b, then clearly $x \in C(X)$ and $x'MP = b'X'(I - M_{MP})MP = 0$ because $(I - M_{MP})MP = 0$ □

Note also that $C(X) \cap C(MP)^{\perp} = C(X) \cap C(P)^{\perp}$.

EXAMPLE 3.3.4. To illustrate these ideas, consider testing $H_0 : \alpha_1 - \alpha_3 = 0$ in Example 3.3.1. The constraint can be written

$$0 = \alpha_1 - \alpha_3 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \frac{-1}{2}, \frac{-1}{2}\right) \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix},$$

so $P' = (1/3, 1/3, 1/3, 0, -1/2, -1/2)$. We need $C(X) \cap C(MP)^{\perp}$. Note that vectors in $C(X)$ have the form $(a, a, a, b, c, c)'$ for any a, b, c, so $P = MP$. Vectors in $C(MP)^{\perp}$ are $v = (v_{11}, v_{12}, v_{13}, v_{21}, v_{31}, v_{32})'$ with $P'v = \bar{v}_{1\cdot} - \bar{v}_{3\cdot} = 0$. Vectors in $C(X) \cap C(MP)^{\perp}$ have the first three elements identical, the last two elements identical, and the average of the first three equal to the average of the last two, i.e., they have the form $(a, a, a, b, a, a)'$. A spanning set for this space is given by the columns of

$$X_0 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

As seen earlier, another choice for P is $(1, 0, 0, 0, -1, 0)'$. Using M from Exercise 1.5.8b, this choice for P also leads to $MP = (1/3, 1/3, 1/3, 0, -1/2, -1/2)'$. To compute $(I - M_{MP})X$, observe that

$$M_{MP} = \frac{1}{(1/3)+(1/2)} \begin{bmatrix} 1/9 & 1/9 & 1/9 & 0 & -1/6 & -1/6 \\ 1/9 & 1/9 & 1/9 & 0 & -1/6 & -1/6 \\ 1/9 & 1/9 & 1/9 & 0 & -1/6 & -1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1/6 & -1/6 & -1/6 & 0 & 1/4 & 1/4 \\ -1/6 & -1/6 & -1/6 & 0 & 1/4 & 1/4 \end{bmatrix}$$

$$= \frac{1}{5} \begin{bmatrix} 2/3 & 2/3 & 2/3 & 0 & -1 & -1 \\ 2/3 & 2/3 & 2/3 & 0 & -1 & -1 \\ 2/3 & 2/3 & 2/3 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 3/2 & 3/2 \\ -1 & -1 & -1 & 0 & 3/2 & 3/2 \end{bmatrix}.$$

Then

$$(I - M_{MP})X = X - M_{MP}X$$

$$= X - \frac{1}{5} \begin{bmatrix} 0 & 2 & 0 & -2 \\ 0 & 2 & 0 & -2 \\ 0 & 2 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & -3 & 0 & 3 \\ 0 & -3 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 3/5 & 0 & 2/5 \\ 1 & 3/5 & 0 & 2/5 \\ 1 & 3/5 & 0 & 2/5 \\ 1 & 0 & 1 & 0 \\ 1 & 3/5 & 0 & 2/5 \\ 1 & 3/5 & 0 & 2/5 \end{bmatrix},$$

which has the same column space as X_0 given earlier.

We have reduced the problem of finding X_0 to that of finding $C(X) \cap C(MP)^{\perp}$, which is just the orthogonal complement of $C(MP)$ with respect to $C(X)$. By Theorem B.48, $C(X) \cap C(MP)^{\perp} = C(M - M_{MP})$, so $M - M_{MP}$ is another valid choice for X_0. For Example 3.3.1 with $H_0 : \alpha_1 - \alpha_3 = 0$, M was given in Exercise 1.5.8b and M_{MP} was given earlier, so

$$M - M_{MP} = \begin{bmatrix} 1/5 & 1/5 & 1/5 & 0 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 0 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 0 & 1/5 & 1/5 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 0 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 0 & 1/5 & 1/5 \end{bmatrix}.$$

This matrix has the same column space as the other choices of X_0 that have been given.

For $\Lambda'\beta$ estimable, we now rewrite the test statistic in (5) in terms of Λ and $\hat{\beta}$. First, we wish to show that $r(\Lambda) = r(M_{MP})$. It suffices to show that $r(\Lambda) = r(MP)$. Writing $\Lambda = X'P$, we see that for any vector b, $X'Pb = 0$ if and only if $Pb \perp C(X)$, which occurs if and only if $MPb = 0$. It follows that $C(P'X)^{\perp} = C(P'M)^{\perp}$ so that $C(P'X) = C(P'M)$, $r(P'X) = r(P'M)$, and $r(\Lambda) = r(X'P) = r(MP)$.

Now rewrite the quadratic form $Y'M_{MP}Y$. Recall that since $X\hat{\beta} = MY$, we have $\Lambda'\hat{\beta} = P'X\hat{\beta} = P'MY$. Substitution gives

$$\begin{aligned} Y'M_{MP}Y &= Y'MP(P'MP)^{-}P'MY \\ &= \hat{\beta}'\Lambda(P'X(X'X)^{-}X'P)^{-}\Lambda'\hat{\beta} \\ &= \hat{\beta}'\Lambda[\Lambda'(X'X)^{-}\Lambda]^{-}\Lambda'\hat{\beta}. \end{aligned}$$

The test statistic in (5) becomes

$$\frac{\hat{\beta}'\Lambda[\Lambda'(X'X)^{-}\Lambda]^{-}\Lambda'\hat{\beta}/r(\Lambda)}{MSE}.$$

A similar argument shows that the noncentrality parameter in (5) can be written as $\beta'\Lambda[\Lambda'(X'X)^{-}\Lambda]^{-}\Lambda'\beta/2\sigma^2$. The test statistic consists of three main parts: MSE, $\Lambda'\hat{\beta}$, and the generalized inverse of $\Lambda'(X'X)^{-}\Lambda$. Note that $\sigma^2\Lambda'(X'X)^{-}\Lambda = \mathrm{Cov}(\Lambda'\hat{\beta})$. *These facts give an alternative method of deriving tests. One can simply find the estimate $\Lambda'\hat{\beta}$, the covariance matrix of the estimate, and the MSE.*

For a single degree of freedom hypothesis $H_0 : \lambda'\beta = 0$, the numerator takes the especially nice form

$$\hat{\beta}'\lambda[\lambda'(X'X)^{-}\lambda]^{-1}\lambda'\hat{\beta} = (\lambda'\hat{\beta})^2/[\lambda'(X'X)^{-}\lambda];$$

so the F test becomes: reject $H_0 : \lambda'\beta = 0$ if

$$\frac{(\lambda'\hat{\beta})^2}{MSE[\lambda'(X'X)^{-}\lambda]} > F(1-\alpha, 1, dfE),$$

which is just the square of the t test that could be derived from the sampling distributions of the least squares estimate and the MSE, cf. Exercise 2.1.

Definition 3.3.5. The condition $\mathrm{E}(Y) \perp C(MP)$ is called the *constraint* on the model caused (imposed) by $\Lambda'\beta = 0$, where $\Lambda' = P'X$. As a shorthand, we will call $C(MP)$ the constraint caused by $\Lambda'\beta = 0$. If $\mathscr{M} \subset C(X)$ and $C(MP) \subset \mathscr{M}$, we say that $C(MP)$ is the constraint on $\mathscr{M}$ caused by $\Lambda'\beta = 0$. If $\Lambda'\beta = 0$ puts a constraint on $\mathscr{M}$, we say that $\Lambda'\beta = 0$ is an hypothesis in $\mathscr{M}$.

Exercise 3.5 Show that a necessary and sufficient condition for $\rho_1'X\beta = 0$ and $\rho_2'X\beta = 0$ to determine orthogonal constraints on the model is that $\rho_1'M\rho_2 = 0$.

Theoretical Complements

If, rather than testing the constraint $\Lambda'\beta = 0$, our desire is to estimate β subject to the constraint, simply estimate γ in model (3) and use $\hat{\beta} = U\hat{\gamma}$. In this *constrained estimation*, the estimates automatically satisfy $\Lambda'\hat{\beta} = 0$. Moreover, estimable functions $\Gamma'\beta = Q'X\beta$ are equivalent to $\Gamma'U\gamma = Q'XU\gamma$, and optimal estimates of γ are transformed into optimal estimates of β.

We now examine the implications of testing $\Lambda'\beta = 0$ when $\Lambda'\beta$ is not estimable. Recall that we began this section by finding the reduced model associated with such a constraint, so we already have a general method for performing such tests.

The first key result is that in defining a linear constraint there is no reason to use anything but estimable functions, because only estimable functions induce a real constraint on $C(X)$. Theorem 3.3.6 identifies the *estimable part* of $\Lambda'\beta$, say $\Lambda_0'\beta$, and implies that $\Lambda_0'\beta = 0$ gives the same reduced model as $\Lambda'\beta = 0$. Λ_0 is a matrix chosen so that $C(\Lambda) \cap C(X') = C(\Lambda_0)$. With such a choice, $\Lambda'\beta = 0$ implies that $\Lambda_0'\beta = 0$ but $\Lambda_0'\beta$ is estimable because $C(\Lambda_0) \subset C(X')$, so $\Lambda_0' = P_0'X$ for some P_0.

Theorem 3.3.6. If $C(\Lambda) \cap C(X') = C(\Lambda_0)$ and $C(U_0) = C(\Lambda_0)^{\perp}$, then $C(XU) = C(XU_0)$. Thus $\Lambda'\beta = 0$ and $\Lambda_0'\beta = 0$ induce the same reduced model.

PROOF. $C(\Lambda_0) \subset C(\Lambda)$, so $C(U) = C(\Lambda)^{\perp} \subset C(\Lambda_0)^{\perp} = C(U_0)$ and $C(XU) \subset C(XU_0)$.

To complete the proof, we show that there cannot be any vectors in $C(XU_0)$ that are not in $C(XU)$. In particular, we show that there are no nontrivial vectors in $C(XU_0)$ that are orthogonal to $C(XU)$, i.e., if $v \in C(XU)^{\perp}_{C(XU_0)}$ then $v = 0$. If $v \in C(XU)^{\perp}_{C(XU_0)}$, then $v'XU = 0$, so $X'v \perp C(U)$ and $X'v \in C(\Lambda)$. But also note that $X'v \in C(X')$, so $X'v \in C(\Lambda) \cap C(X') = C(\Lambda_0)$. This implies that $X'v \perp C(U_0)$, so $v \perp C(XU_0)$. We have shown that the vector v which, by assumption, is in $C(XU_0)$, is also orthogonal to $C(XU_0)$. The only such vector is the 0 vector. □

Nontrivial estimable constraints always induce a real constraint on the column space.

Proposition 3.3.7. If $\Lambda'\beta$ is estimable and $\Lambda \neq 0$, then $\Lambda'\beta = 0$ implies that $C(XU) \neq C(X)$.

PROOF. With $\Lambda' = P'X$, the definition of U gives $0 = \Lambda'U = P'XU = P'MXU$, so $C(XU) \perp C(MP)$. Both are subspaces of $C(X)$; therefore if $C(MP) \neq \{0\}$, we have $C(X) \neq C(XU)$. However, $P'MX = \Lambda' \neq 0$, so $C(MP)$ is not orthogonal to $C(X)$ and $C(MP) \neq \{0\}$. □

Note that Proposition 3.3.7 also implies that whenever the estimable part Λ_0 is different from 0, there is always a real constraint on the column space.

Corollary 3.3.8 establishes that $\Lambda'\beta$ has no estimable part if and only if the constraint does not affect the model. If the constraint does not affect the model, it merely defines a reparameterization, in other words, it merely specifies arbitrary side conditions. The corollary follows from the observation made about Λ_0 after Proposition 3.3.7 and taking $\Lambda_0 = 0$ in Theorem 3.3.6.

Corollary 3.3.8. $C(\Lambda) \cap C(X') = \{0\}$ if and only if $C(XU) = C(X)$.

In particular, if $\Lambda'\beta$ is not estimable, we can obtain the numerator sum of squares for testing $\Lambda'\beta = 0$ either by finding $X_0 = XU$ directly and using it to get $M - M_0$, or by finding Λ_0, writing $\Lambda_0' = P_0'X$, and using M_{MP_0}. But as noted earlier, there is no reason to have $\Lambda'\beta$ not estimable.

3.3.1 A Generalized Test Procedure

We now consider hypotheses of the form $\Lambda'\beta = d$ where $d \in C(\Lambda')$ so that the equation $\Lambda'\beta = d$ is solvable. Let b be such a solution. Note that

$$\Lambda'\beta = \Lambda'b = d$$

if and only if

$$\Lambda'(\beta - b) = 0$$

if and only if

$$(\beta - b) \perp C(\Lambda).$$

Again picking a matrix U such that

$$C(U) = C(\Lambda)^{\perp},$$

$\Lambda'(\beta - b) = 0$ if and only if

$$(\beta - b) \in C(U),$$

which occurs if and only if for some vector γ

$$(\beta - b) = U\gamma.$$

Multiplying both sides by X gives

$$X\beta - Xb = XU\gamma$$

or

$$X\beta = XU\gamma + Xb.$$

We can now substitute this into the linear model to get the reduced model

$$Y = XU\gamma + Xb + e,$$

or, letting $X_0 \equiv XU$,

$$Y = X_0\gamma + Xb + e. \tag{6}$$

Recall that b is a vector we can find, so Xb is a known (offset) vector.

The analysis for reduced models such as (6) was developed in Section 2. For nonestimable linear hypotheses, use that theory directly. If $\Lambda' = P'X$, then $C(X_0)^{\perp}_{C(X)} = C(MP)$ and the test statistic is easily seen to be

$$\frac{(Y-Xb)'M_{MP}(Y-Xb)/r(M_{MP})}{(Y-Xb)'(I-M)(Y-Xb)/r(I-M)}.$$

Note that $\Lambda'\beta = d$ imposes the constraint $\mathrm{E}(Y-Xb) \perp C(MP)$, so once again we could refer to $C(MP)$ as the constraint imposed by the hypothesis.

We did not specify the solution b to $\Lambda'\beta = d$ that should be used. Fortunately, for $\Lambda'\beta$ estimable, the test does not depend on the choice of b. As mentioned in the previous section, $(Y-Xb)'(I-M)(Y-Xb) = Y'(I-M)Y$, so the denominator of the test is just the MSE and does not depend on b. The numerator term $(Y-Xb)'M_{MP}(Y-Xb)$ equals $(\Lambda'\hat{\beta}-d)'[\Lambda'(X'X)^-\Lambda]^-(\Lambda'\hat{\beta}-d)$. The test statistic can be written as

$$\frac{(\Lambda'\hat{\beta}-d)'[\Lambda'(X'X)^-\Lambda]^-(\Lambda'\hat{\beta}-d)/r(\Lambda)}{MSE}.$$

For $\Lambda'\beta$ estimable, the linear model $Y = X_0\gamma + Xb + e$ implies that $\Lambda'\beta = d$, but for nonestimable linear constraints, there are infinitely many constraints that result in the same reduced model. (If you think of nonestimable linear constraints as including arbitrary side conditions, that is not surprising.) In particular, if $\Lambda'\beta = d$, the same reduced model results if we take $\Lambda'\beta = d_0$ where $d_0 = d + \Lambda'v$ and $v \perp C(X')$. Note that, in this construction, if $\Lambda'\beta$ is estimable, $d_0 = d$ for any v.

We now present an application of this testing procedure. The results are given without justification, but they should seem similar to results from a statistical methods course.

EXAMPLE 3.3.9. Consider the balanced two-way ANOVA without interaction model

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

$i = 1,\ldots,a$, $j = 1,\ldots,b$, $k = 1,\ldots,N$. (The analysis for this model is presented in Section 7.1.) We examine the test of the null hypothesis

$$H_0 : \sum_{i=1}^{a} \lambda_i\alpha_i = 4 \quad \text{and} \quad \sum_{j=1}^{b} \gamma_j\eta_j = 7,$$

where $\sum_{i=1}^{a} \lambda_i = 0 = \sum_{j=1}^{b} \gamma_j$. The hypothesis is simultaneously specifying the values of a contrast in the α_is and a contrast in the η_js.

In terms of the model $Y = X\beta + e$, we have

$$\beta = [\mu, \alpha_1, \ldots, \alpha_a, \eta_1, \ldots, \eta_b]'$$

$$\Lambda' = \begin{bmatrix} 0 & \lambda_1 & \cdots & \lambda_a & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \gamma_1 & \cdots & \gamma_b \end{bmatrix}$$

$$d = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

$$\Lambda'\hat{\beta} = \begin{bmatrix} \sum_{i=1}^{a} \lambda_i \bar{y}_{i\cdot\cdot} \\ \sum_{j=1}^{b} \gamma_j \bar{y}_{\cdot j\cdot} \end{bmatrix}$$

$$\mathrm{Cov}(\Lambda'\hat{\beta})/\sigma^2 = \Lambda'(X'X)^-\Lambda = \begin{bmatrix} \sum_{i=1}^{a} \lambda_i^2/bN & 0 \\ 0 & \sum_{j=1}^{b} \gamma_j^2/aN \end{bmatrix}.$$

The diagonal elements of the covariance matrix are just the variances of the estimated contrasts. The off-diagonal elements are zero because this is a balanced two-way ANOVA, hence the estimates of the α contrast and the η contrast are independent. We will see in Chapter 7 that these contrasts define orthogonal constraints in the sense of Definition 3.3.5, so they are often referred to as being orthogonal parameters.

There are two linearly independent contrasts being tested, so $r(\Lambda) = 2$. The test statistic is

$$\frac{1}{2MSE} \left[\sum_{i=1}^{a} \lambda_i \bar{y}_{i\cdot\cdot} - 4 \quad \sum_{j=1}^{b} \gamma_j \bar{y}_{\cdot j\cdot} - 7 \right] \times \begin{bmatrix} \frac{bN}{\sum_{i=1}^{a} \lambda_i^2} & 0 \\ 0 & \frac{aN}{\sum_{j=1}^{b} \gamma_j^2} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{a} \lambda_i \bar{y}_{i\cdot\cdot} - 4 \\ \sum_{j=1}^{b} \gamma_j \bar{y}_{\cdot j\cdot} - 7 \end{bmatrix}$$

or

$$\frac{1}{2MSE} \left[\frac{\left(\sum_{i=1}^{a} \lambda_i \bar{y}_{i\cdot\cdot} - 4\right)^2}{\sum_{i=1}^{a} \lambda_i^2/bN} + \frac{\left(\sum_{j=1}^{b} \gamma_j \bar{y}_{\cdot j\cdot} - 7\right)^2}{\sum_{j=1}^{b} \gamma_j^2/aN} \right].$$

Note that the term $\left(\sum_{i=1}^{a} \lambda_i \bar{y}_{i\cdot\cdot} - 4\right)^2 / \left(\sum_{i=1}^{a} \lambda_i^2/bN\right)$ is, except for subtracting the 4, the sum of squares for testing $\sum_{i=1}^{a} \lambda_i \alpha_i = 0$ We are subtracting the 4 because we are testing $\sum_{i=1}^{a} \lambda_i \alpha_i = 4$. Similarly, we have a term that is very similar to the sum of squares for testing $\sum_{j=1}^{b} \gamma_j \eta_j = 0$. The test statistic takes the average of these sums of squares and divides by the MSE. The test is then defined by reference to an $F(2, dfE, 0)$ distribution.

3.3.2 Testing an Unusual Class of Hypotheses

Occasionally, a valid linear hypothesis $\Lambda'\beta = d$ is considered where d is not completely known but involves other parameters. (This is the linear structure involved in creating a logistic regression model from a log-linear model.) For $\Lambda'\beta = d$ to give a valid linear hypothesis, $\Lambda'\beta = d$ must put a restriction on $C(X)$, so that when the hypothesis is true, $\mathrm{E}(Y)$ lies in some subspace of $C(X)$.

Let X_1 be such that $C(X_1) \subset C(X)$ and consider an hypothesis

$$P'X\beta = P'X_1\delta$$

for some parameter vector δ. We seek an appropriate reduced model for such an hypothesis.

Note that the hypothesis occurs if and only if

$$P'M(X\beta - X_1\delta) = 0,$$

which occurs if and only if

$$(X\beta - X_1\delta) \perp C(MP),$$

which occurs if and only if

$$(X\beta - X_1\delta) \in C(MP)^{\perp}_{C(X)}.$$

As discussed earlier in this section, we choose X_0 so that $C(MP)^{\perp}_{C(X)} = C(X_0)$. The choice of X_0 does not depend on X_1. Using X_0, the hypothesis occurs if and only if for some γ

$$(X\beta - X_1\delta) = X_0\gamma.$$

Rewriting these terms, we see that

$$X\beta = X_0\gamma + X_1\delta$$

which is the mean structure for the reduced model. In other words, assuming the null hypothesis is equivalent to assuming a reduced model

$$Y = X_0\gamma + X_1\delta + e.$$

To illustrate, consider a linear model for pairs of observations (y_{1j}, y_{2j}), $j = 1, \ldots, N$. Write $Y = (y_{11}, \ldots, y_{1N}, y_{21}, \ldots, y_{2N})'$. Initially, we will impose no structure on the means so that $\mathrm{E}(y_{ij}) = \mu_{ij}$. We are going to consider an hypothesis for the differences between the pairs,

$$\mu_{1j} - \mu_{2j} = z_j'\delta$$

for some known predictor vector z_j. Of course we could just fit a linear model to the differences $y_{1j} - y_{2j}$, but we want to think about comparing such a model to models that are not based on the differences.

The conditions just specified correspond to a linear model $Y = X\beta + e$ in which $X = I_{2N}$ and $\beta = (\mu_{11}, \ldots, \mu_{2N})'$. Write

$$P = \begin{bmatrix} I_N \\ -I_N \end{bmatrix} \quad \text{and} \quad X_1 = \begin{bmatrix} Z \\ 0 \end{bmatrix}$$

where $Z' = [z_1, \ldots, z_N]$. Then the hypothesis for the differences can be specified as

$$P'X\beta = P'X_1\delta = Z\delta.$$

Finally, it is not difficult to see that a valid choice of X_0 is

$$X_0 = \begin{bmatrix} I_N \\ I_N \end{bmatrix}.$$

It follows that, under the reduced model

$$X\beta \equiv I\beta = \begin{bmatrix} I & Z \\ I & 0 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix}$$

or that the reduced model is

$$Y = \begin{bmatrix} I & Z \\ I & 0 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix} + e.$$

This relationship is of particular importance in the analysis of frequency data. The model $y_{ij} = \mu_{ij} + e_{ij}$ is analogous to a saturated log-linear model. An hypothesis $\mu_{1j} - \mu_{2j} = \alpha_0 + \alpha_1 t_j \equiv z_j'\delta$ is analogous to the hypothesis that a simple linear logit model holds. We have found the vector space such that restricting the log-linear model to that space gives the logit model, see also Christensen (1997).

Exercise 3.6 In testing a reduced model $Y = X_0\gamma + e$ against a full model $Y = X\beta + e$, what linear parametric function of the parameters is being tested?

3.4 Discussion

The reason that we considered testing models first and then discussed testing parametric functions by showing them to be changes in models is because, in *general,* only model testing is ever performed. This is not to say that parametric functions are not tested as such, but that parametric functions are only tested in special cases. In particular, parametric functions can easily be tested in balanced ANOVA problems and one-way ANOVAs. Multifactor ANOVA designs with unequal numbers of observations in the treatment cells, as illustrated in Example 3.2.2, are best analyzed by considering alternative models. Even in regression models, where all the parameters are estimable, it is often more enlightening to think in terms of model selection. Of course, in regression there is a special relationship between the parameters and the model matrix. For the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{ip-1} + e$, the model matrix can be written as $X = [J, X_1, \ldots, X_{p-1}]$, where $X_j = [x_{1j}, x_{2j}, \ldots, x_{nj}]'$. The test of $H_0 : \beta_j = 0$ is obtained by just leaving X_j out of the model matrix.

Another advantage of the method of testing models is that it is often easy in simple but nontrivial cases to see immediately what new model is generated by a null hypothesis. This was illustrated in Examples 3.2.0 and 3.2.3.

EXAMPLE 3.4.1. *One-Way ANOVA.*
Consider the model $y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1,2,3$, $j = 1,\ldots,N_i$, $N_1 = N_3 = 3$, $N_2 = 2$. In matrix terms this is

$$Y = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + e.$$

Let the null hypothesis be $\alpha_1 = \mu + 2\alpha_2$. Writing $X = [J, X_1, X_2, X_3]$ and

$$\mathrm{E}(Y) = X\beta = \mu J + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3,$$

the reduced model is easily found by substituting $\mu + 2\alpha_2$ for α_1 which leads to

$$\mathrm{E}(Y) = \mu(J + X_1) + \alpha_2(2X_1 + X_2) + \alpha_3 X_3.$$

This gives the reduced model

$$Y = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 2 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} + e.$$

For Examples 3.2.0, 3.2.3, and 3.5.1, it would be considerable work to go through the procedure developed in Section 3 to test the hypotheses. In fairness, it should be added that for these special cases, there is no need to go through the general procedures of Section 3 to get the tests (assuming that you get the necessary computer output for the regression problem).

3.5 Testing Single Degrees of Freedom in a Given Subspace

Consider a two-way ANOVA model $y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}$. Suppose we want to look at contrasts in the α_is and η_js. For analyzing a balanced two-way ANOVA it would be very convenient if estimates and tests for contrasts in the α_is, say, could be based on the projection operator associated with dropping the α_is out of the one-way ANOVA model $y_{ijk} = \mu + \alpha_i + e_{ijk}$ rather than the projection operator for the two-way ANOVA model. One convenience is that the projection operator for the one-way model turns out to be much simpler than the projection operator for the two-way model. A second convenience is that orthogonality of the projection operators for dropping the α_is and η_js in the balanced two-way model leads to independence between estimates of contrasts in the α_is and η_js. We would also like to establish that orthogonal contrasts (contrasts that define orthogonal constraints) in the α_is, say, depend only on the projection operator for dropping the α_is in the one-way model.

With these ultimate goals in mind, we now examine, in general, estimates, tests, and orthogonality relationships between single degree of freedom hypotheses that put a constraint on a particular subspace.

Consider a perpendicular projection operator M_* used in the numerator of a test statistic. In the situation of testing a model $Y = X\beta + e$ against a reduced model $Y = X_0\gamma + e$ with $C(X_0) \subset C(X)$, if M and M_0 are the perpendicular projection operators onto $C(X)$ and $C(X_0)$, respectively, then $M_* = M - M_0$. For testing the estimable parametric hypothesis $\Lambda'\beta = 0$, if $\Lambda' = P'X$, then $M_* = M_{MP}$.

We want to examine the problem of testing a single degree of freedom hypothesis in $C(M_*)$. Let $\lambda' = \rho'X$. Then, by Definition 3.3.2, $\lambda'\beta = 0$ puts a constraint on $C(M_*)$ if and only if $M\rho \in C(M_*)$. If $M\rho \in C(M_*)$, then $M\rho = M_*M\rho = M_*\rho$ because $MM_* = M_*$. It follows that the estimate of $\lambda'\beta$ is $\rho'M_*Y$ because $\rho'M_*Y = \rho'MY$. From Section 3, the test statistic for $H_0 : \lambda'\beta = 0$ is

$$\frac{Y'M_*\rho(\rho'M_*\rho)^{-1}\rho'M_*Y}{MSE} = \frac{(\rho'M_*Y)^2/\rho'M_*\rho}{MSE},$$

where $MSE = Y'(I-M)Y/r(I-M)$ and $r(M_*\rho(\rho'M_*\rho)^{-1}\rho'M_*) = 1$.

Let $\lambda_1' = \rho_1'X$ and $\lambda_2' = \rho_2'X$, and let the hypotheses $\lambda_1'\beta = 0$ and $\lambda_2'\beta = 0$ define orthogonal constraints on the model. The constraints are, respectively, $\mathrm{E}(Y) \perp M\rho_1$ and $\mathrm{E}(Y) \perp M\rho_2$. These constraints are said to be orthogonal if the vectors $M\rho_1$ and $M\rho_2$ are orthogonal. This occurs if and only if $\rho_1'M\rho_2 = 0$. If $\lambda_1'\beta = 0$ and $\lambda_2'\beta = 0$ both put constraints on $C(M_*)$, then orthogonality is equivalent to $0 = \rho_1'M\rho_2 = \rho_1'M_*\rho_2$.

We have now shown that for any estimable functions that put constraints on $C(M_*)$, estimates, tests, and finding orthogonal constraints in $C(M_*)$ require only the projection operator M_* and the MSE.

Exercise 3.7 Show that $\rho'MY = \rho'[M\rho(\rho'M\rho)^{-}\rho'M]Y$ so that to estimate $\rho'X\beta$, one only needs the perpendicular projection of Y onto $C(M\rho)$.

3.6 Breaking a Sum of Squares into Independent Components

We now present a general theory that includes, as special cases, the breaking down of the treatment sum of squares in a one-way ANOVA into sums of squares for orthogonal contrasts and the breaking of the sum of squares for the model into independent sums of squares as in an ANOVA table. This is an important device in statistical analyses.

Frequently, a reduced model matrix X_0 is a submatrix of X. This is true for the initial hypotheses considered in both cases of Example 3.2.0 and for Example 3.2.2. If we can write $X = [X_0, X_1]$, it is convenient to write $SSR(X_1|X_0) \equiv Y'(M - M_0)Y$. $SSR(X_1|X_0)$ is called the sum of squares for regressing X_1 after X_0. We will also write $SSR(X) \equiv Y'MY$, the sum of squares for regressing on X. Similarly, $SSR(X_0) \equiv Y'M_0Y$. The $SSR(\cdot)$ notation is one way of identifying sums of squares for tests. Other notations exist, and one alternative will soon be introduced. Note that $SSR(X) = SSR(X_0) + SSR(X_1|X_0)$, which constitutes a breakdown of $SSR(X)$ into two parts. If $e \sim N(0, \sigma^2 I)$, these two parts are independent.

We begin with a general theory and conclude with a discussion of breaking down the sums of squares in a two-way ANOVA model $y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}$. The projection operators used in the numerator sums of squares for dropping the α_is and η_js are orthogonal if and only if the numerator sum of squares for dropping the η_js out of the two-way model is the same as the numerator sum of squares for dropping the η_js out of the one-way ANOVA model $y_{ijk} = \mu + \eta_j + e_{ijk}$.

3.6.1 *General Theory*

We now present a general theory that is based on finding an orthonormal basis for a subspace of the estimation space. (This subspace could be the entire estimation space.) We discuss two methods of doing this. The first is a direct method involving identifying the subspace and choosing an orthonormal basis. The second method determines an orthonormal basis indirectly by examining single degree of freedom hypotheses and the constraints imposed by those hypotheses.

Our general model is $Y = X\beta + e$ with M the perpendicular projection operator onto $C(X)$. Let M_* be any perpendicular projection operator with $C(M_*) \subset C(X)$. Then M_* defines a test statistic

$$\frac{Y'M_*Y/r(M_*)}{Y'(I-M)Y/r(I-M)}$$

for testing the reduced model, say, $Y = (M - M_*)\gamma + e$. If $r(M_*) = r$, then we will show that we can break the sum of squares based on M_* (i.e., $Y'M_*Y$) into as many as r independent sums of squares whose sum will equal $Y'M_*Y$. By using M_* in the numerator of the test, we are testing whether the subspace $C(M_*)$ is adding anything to the predictive (estimative) ability of the model. What we have done is break $C(X)$

into two orthogonal parts, $C(M_*)$ and $C(M - M_*)$. In this case, $C(M - M_*)$ is the estimation space under H_0 and we can call $C(M_*)$ the *test space*. $C(M_*)$ is a space that will contain only error if H_0 is true but which is part of the estimation space under the full model. Note that the error space under H_0 is $C(I - (M - M_*))$, but $I - (M - M_*) = (I - M) + M_*$ so that $C(I - M)$ is part of the error space under both models.

We now break $C(M_*)$ into r orthogonal subspaces. Take an orthonormal basis for $C(M_*)$, say $R_1, R_2, \ldots, R_r$. Note that, using Gram–Schmidt, R_1 can be any normalized vector in $C(M_*)$. It is the statistician's choice. R_2 can then be any normalized vector in $C(M_*)$ orthogonal to R_1, etc. Let $R = [R_1, R_2, \ldots, R_r]$, then as in Theorem B.35,

$$M_* = RR' = [R_1, \ldots, R_r] \begin{bmatrix} R_1' \\ \vdots \\ R_r' \end{bmatrix} = \sum_{i=1}^{r} R_i R_i'.$$

Let $M_i = R_i R_i'$, then M_i is a perpendicular projection operator in its own right and $M_i M_j = 0$ for $i \neq j$ because of the orthogonality of the R_is.

The goal of this section is to break up the sum of squares into independent components. By Theorem 1.3.7, the sums of squares $Y'M_iY$ and $Y'M_jY$ are independent for any $i \neq j$ because $M_i M_j = 0$. Also, $Y'M_*Y = \sum_{i=1}^{r} Y'M_iY$ simply because $M_* = \sum_{i=1}^{r} M_i$. Moreover, since $r(M_i) = 1$,

$$\frac{Y'M_iY}{Y'(I-M)Y/r(I-M)} \sim F(1, r(I-M), \beta'X'M_iX\beta/2\sigma^2).$$

In a one-way ANOVA, $Y'M_*Y$ corresponds to the treatment sum of squares while the $Y'M_iY$s correspond to the sums of squares for a set of orthogonal contrasts, cf. Example 3.6.2.

We now consider the correspondence between the hypothesis tested using $Y'M_*Y$ and those tested using the $Y'M_iY$s. Because M_* and the M_is are nonnegative definite,

$$0 = \beta'X'M_*X\beta = \sum_{i=1}^{r} \beta'X'M_iX\beta$$

if and only if $\beta'X'M_iX\beta = 0$ for all i if and only if $R_i'X\beta = 0$ for all i. Thus, the null hypothesis that corresponds to the test based on M_* is true if and only if the null hypotheses $R_i'X\beta = 0$ corresponding to all the M_is are true. Equivalently, if the null hypothesis corresponding to M_* is not true, we have

$$0 < \beta'X'M_*X\beta = \sum_{i=1}^{r} \beta'X'M_iX\beta.$$

Again, since M_* and the M_is are nonnegative definite, this occurs if and only if at least one of the terms $\beta'X'M_iX\beta$ is greater than zero. Thus the null hypothesis corresponding to M_* is not true if and only if at least one of the hypotheses corresponding to the M_is is not true. Thinking in terms of a one-way ANOVA, these results correspond to stating that 1) the hypothesis of no treatment effects is true if and only if all the contrasts in a set of orthogonal contrasts are zero or, equivalently, 2) the hypothesis of no treatment effects is not true if and only if at least one contrast in a set of orthogonal contrasts is not zero.

We have broken $Y'M_*Y$ into r independent parts. It is easy to see how to break it into less than r parts. Suppose $r = 7$. We can break $Y'M_*Y$ into three parts by looking at projections onto only three subspaces. For example, $Y'M_*Y = Y'(M_1 + M_3 + M_6)Y + Y'(M_2 + M_7)Y + Y'(M_4 + M_5)Y$, where we have used three projection operators $M_1 + M_3 + M_6$, $M_2 + M_7$, and $M_4 + M_5$. Note that these three projection operators are orthogonal, so the sums of squares are independent. By properly choosing R, an ANOVA table can be developed using this idea.

EXAMPLE 3.6.1. *One-Way ANOVA.*
In this example we examine breaking up the treatment sum of squares in a one-way ANOVA. Consider the model $y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1,2,3$, $j = 1,2,3$. In matrix terms this is

$$Y = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + e. \qquad (1)$$

Denote the model matrix $X = [J, X_1, X_2, X_3]$. To test $H_0 : \alpha_1 = \alpha_2 = \alpha_3$, the reduced model is clearly

$$Y = J\mu + e.$$

The projection operator for the test is $M_* = M - [1/n]JJ'$. The test space is $C(M_*) = C(M - [1/n]JJ')$, i.e., the test space is the set of all vectors in $C(X)$ that are orthogonal to a column of ones. The test space can be obtained by using Gram–Schmidt to remove the effect of J from the last three columns of the model matrix, that is, $C(M_*)$ is spanned by the columns of

$$\begin{bmatrix} 2 & -1 & -1 \\ 2 & -1 & -1 \\ 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \\ -1 & -1 & 2 \\ -1 & -1 & 2 \end{bmatrix},$$

which is a rank 2 matrix. The statistician is free to choose R_1 within $C(M_*)$. R_1 could be a normalized version of

$$\begin{bmatrix} 2 \\ 2 \\ 2 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ -1 \\ 2 \\ 2 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -2 \\ -2 \\ -2 \end{bmatrix},$$

which was chosen as $X_1 + X_2$ with the effect of J removed. R_2 must be the only normalized vector left in $C(M_*)$ that is orthogonal to R_1. R_2 is a normalized version of $[1,1,1,-1,-1,-1,0,0,0]'$. The sum of squares for testing $H_0 : \alpha_1 = \alpha_2 = \alpha_3$ is $Y'R_1R_1'Y + Y'R_2R_2'Y$.

Using the specified form of R_1, $Y'M_1Y$ is the numerator sum of squares for testing

$$0 = R_1'X\beta \propto (0,3,3,-6) \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = 6\left(\frac{\alpha_1+\alpha_2}{2} - \alpha_3\right).$$

Similarly,

$$R_2'X\beta \propto \alpha_1 - \alpha_2.$$

R_1 was chosen so that $Y'R_2R_2'Y$ would be the sum of squares for testing $H_0 : \alpha_1 = \alpha_2$.

The discussion thus far has concerned itself with directly choosing an orthonormal basis for $C(M_*)$. An equivalent approach to the problem of finding an orthogonal breakdown is in terms of single degree of freedom hypotheses $\lambda'\beta = 0$.

If we choose any r single degree of freedom hypotheses $\lambda_1'\beta = \cdots = \lambda_r'\beta = 0$ with $\rho_k'X = \lambda_k'$, $M\rho_k \in C(M_*)$, and $\rho_k'M\rho_h = 0$ for all $k \neq h$, then the vectors $M\rho_k/\sqrt{\rho_k'M\rho_k}$ form an orthonormal basis for $C(M_*)$. The projection operators are $M_k = M\rho_k(\rho_k'M\rho_k)^{-1}\rho_k'M$. The sums of squares for these hypotheses, $Y'M_kY = Y'M\rho_k(\rho_k'M\rho_k)^{-1}\rho_k'MY = (\rho_k'MY)^2/\rho_k'M\rho_k = (\rho_k M_* Y)^2/\rho_k'M_*\rho_k$, form an orthogonal breakdown of $Y'M_*Y$.

As shown in Section 3, the sum of squares for testing $\lambda_k'\beta = 0$ can be found from λ_k, $\hat{\beta}$, and $(X'X)^-$. In many ANOVA problems, the condition $\rho_k'M\rho_h = 0$ can be checked by considering an appropriate function of λ_k and λ_h. It follows that, in many problems, an orthogonal breakdown can be obtained without actually finding the vectors $\rho_1,\ldots,\rho_r$.

EXAMPLE 3.6.2. *One-Way ANOVA.*
Consider the model $y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1,\ldots,t$, $j = 1,\ldots,N_i$. Let $Y'M_*Y$ correspond to the sum of squares for treatments (i.e., the sum of squares for testing

$\alpha_1 = \cdots = \alpha_t$). The hypotheses $\lambda_k'\beta = 0$ correspond to contrasts $c_{k1}\alpha_1 + \cdots + c_{kt}\alpha_t = 0$, where $c_{k1} + \cdots + c_{kt} = 0$. In Chapter 4, it will be shown that contrasts are estimable functions and that any contrast imposes a constraint on the space for testing equality of treatments. In other words, Chapter 4 shows that the $\lambda_k'\beta$s can be contrasts and that if they are contrasts, then $M\rho_k \in C(M_*)$. In Chapter 4 it will also be shown that the condition for orthogonality, $\rho_k' M\rho_h = 0$, reduces to the condition $c_{k1}c_{h1}/N_1 + \cdots + c_{kt}c_{ht}/N_t = 0$. If the contrasts are orthogonal, then the sums of squares for the contrasts add up to the sums of squares for treatments, and the sums of squares for the contrasts are independent.

3.6.2 Two-Way ANOVA

We discuss the technique of breaking up sums of squares as it applies to the two-way ANOVA model of Example 3.2.2. The results really apply to any two-way ANOVA with unequal numbers. The sum of squares for the full model is $Y'MY$ (by definition). We can break this up into three parts, one for fitting the η_js after having fit the α_is and μ, one for fitting the α_is after fitting μ, and one for fitting μ. In Example 3.2.2, the model is $y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}$, $i = 1,2,3$, $j = 1,2,3$. The seven columns of X correspond to the elements of $\beta = [\mu, \alpha_1, \alpha_2, \alpha_3, \eta_1, \eta_2, \eta_3]'$. Let $J = (1, \ldots, 1)'$ be the first column of X. Let X_0 be a matrix consisting of the first four columns of X, those corresponding to μ, α_1, α_2, and α_3. Take M and M_0 corresponding to X and X_0. It is easy to see that $(1/21)JJ'$ is the perpendicular projection operator onto $C(J)$.

Since $J \in C(X_0) \subset C(X)$, we can write, with $n = 21$,

$$Y'MY = Y'\frac{1}{n}JJ'Y + Y'\left(M_0 - \frac{1}{n}JJ'\right)Y + Y'(M - M_0)Y,$$

where $(1/n)JJ'$, $M_0 - (1/n)JJ'$, and $M - M_0$ are all perpendicular projection matrices. Since X_0 is obtained from X by dropping the columns corresponding to the η_js, $Y'(M - M_0)Y$ is the sum of squares used to test the full model against the reduced model with the η_js left out. Recalling our technique of looking at the differences in error sums of squares, we write $Y'(M - M_0)Y \equiv R(\eta|\alpha,\mu)$. $R(\eta|\alpha,\mu)$ is the reduction in (error) sum of squares due to fitting the η_js after fitting μ and the α_is, or, more simply, the sum of squares due to fitting the η_js after the α_is and μ. Similarly, if we wanted to test the model $y_{ijk} = \mu + e_{ijk}$ against $y_{ijk} = \mu + \alpha_i + e_{ijk}$, we would use $Y'(M_0 - [1/n]JJ')Y \equiv R(\alpha|\mu)$, the sum of squares for fitting the α_is after μ. Finally, to test $y_{ijk} = \mu + e_{ijk}$ against $y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}$, we would use $Y'(M - [1/n]JJ')Y \equiv R(\alpha,\eta|\mu)$. Note that $R(\alpha,\eta|\mu) = R(\eta|\alpha,\mu) + R(\alpha|\mu)$.

The notations $SSR(\cdot)$ and $R(\cdot)$ are different notations for essentially the same thing. The $SSR(\cdot)$ notation emphasizes variables and is often used in regression problems. The $R(\cdot)$ notation emphasizes parameters and is frequently used in analysis of variance problems.

Alternatively, we could have chosen to develop the results in this discussion by comparing the model $y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}$ to the model $y_{ijk} = \mu + \eta_j + e_{ijk}$. Then we would have taken X_0 as columns 1, 5, 6, and 7 of the X matrix instead of columns 1, 2, 3, and 4. This would have led to terms such as $R(\eta|\mu)$, $R(\alpha|\eta,\mu)$, and $R(\alpha,\eta|\mu)$. In general, these two analyses *will not* be the same. Typically, $R(\eta|\alpha,\mu) \neq R(\eta|\mu)$ and $R(\alpha|\mu) \neq R(\alpha|\eta,\mu)$. There do exist cases (e.g., balanced two-way ANOVA models) where the order of the analysis has no effect. Specifically, if the columns of the X matrix associated with α and those associated with η are orthogonal after somehow fitting μ, then $R(\eta|\alpha,\mu) = R(\eta|\mu)$ and $R(\alpha|\mu) = R(\alpha|\eta,\mu)$.

As mentioned, the preceding discussion applies to all two-way ANOVA models. We now state precisely the sense in which the columns for α and η need to be orthogonal. Let X_0 be the columns of X associated with μ and the α_is, and let X_1 be the columns of X associated with μ and the η_js. Let M_0 and M_1 be the projection operators onto $C(X_0)$ and $C(X_1)$, respectively. We will show that $R(\eta|\alpha,\mu) = R(\eta|\mu)$ for all Y if and only if $C(M_1 - [1/n]JJ') \perp C(M_0 - [1/n]JJ')$, i.e., $(M_1 - [1/n]JJ')(M_0 - [1/n]JJ') = 0$.

Since $R(\eta|\mu) = Y'(M_1 - [1/n]JJ')Y$ and $R(\eta|\alpha,\mu) = Y'(M - M_0)Y$, it suffices to show the next proposition.

Proposition 3.6.3. In two-way ANOVA, $(M_1 - [1/n]JJ') = (M - M_0)$ if and only if $(M_1 - [1/n]JJ')(M_0 - [1/n]JJ') = 0$.

PROOF. $\Rightarrow$ If $(M_1 - [1/n]JJ') = (M - M_0)$, then

$$(M_1 - [1/n]JJ')(M_0 - [1/n]JJ') = (M - M_0)(M_0 - [1/n]JJ') = 0$$

because $J \in C(M_0) \subset C(M)$.

$\Leftarrow$ To simplify notation, let

$$M_\alpha \equiv (M_0 - [1/n]JJ') \quad \text{and} \quad M_\eta \equiv (M_1 - [1/n]JJ').$$

We know that $M = [1/n]JJ' + M_\alpha + (M - M_0)$. If we could show that $M = [1/n]JJ' + M_\alpha + M_\eta$, we would be done.

$[1/n]JJ' + M_\alpha + M_\eta$ is symmetric and is easily seen to be idempotent since $0 = M_\eta M_\alpha = M_\alpha M_\eta$. It suffices to show that $C[(1/n)JJ' + M_\alpha + M_\eta] = C(X)$. Clearly, $C[(1/n)JJ' + M_\alpha + M_\eta] \subset C(X)$.

Suppose now that $v \in C(X)$. Since $C(M_0) = C(X_0)$ and $C(M_1) = C(X_1)$, if we let $Z = [M_0, M_1]$, then $C(Z) = C(X)$ and $v = Zb = M_0 b_0 + M_1 b_1$. Since $J \in C(X_0)$ and $J \in C(X_1)$, it is easily seen that $M_\alpha M_1 = M_\alpha M_\eta = 0$ and $M_\eta M_0 = 0$. Observe that

$$\begin{aligned}\left[\frac{1}{n}JJ' + M_\alpha + M_\eta\right] v &= [M_0 + M_\eta] M_0 b_0 + [M_1 + M_\alpha] M_1 b_1 \\ &= M_0 b_0 + M_1 b_1 = v,\end{aligned}$$

so $C(X) \subset C[(1/n)JJ' + M_\alpha + M_\eta]$. $\square$

The condition $(M_1 - [1/n]JJ')(M_0 - [1/n]JJ') = 0$ is equivalent to what follows. Using the Gram–Schmidt orthogonalization algorithm, make all the columns corresponding to the αs and ηs orthogonal to J. Now, if the transformed α columns are orthogonal to the transformed η columns, then $R(\eta|\alpha,\mu) = R(\eta|\mu)$. In other words, check the condition $X_0'(I - [1/n]JJ')X_1 = 0$. In particular, this occurs in a balanced two-way ANOVA model, see Section 7.1.

From the symmetry of the problem, it follows that $R(\alpha|\eta,\mu) = R(\alpha|\mu)$ whenever $R(\eta|\alpha,\mu) = R(\eta|\mu)$.

3.7 Confidence Regions

Consider the problem of finding a confidence region for the estimable parametric vector $\Lambda'\beta$. A $(1-\alpha)100\%$ confidence region for $\Lambda'\beta$ consists of all the vectors d that would not be rejected by an α level test of $\Lambda'\beta = d$. That is to say, a $(1-\alpha)100\%$ confidence region for $\Lambda'\beta$ consists of all the vectors d that are consistent with the data and the full model as determined by an α level test of $\Lambda'\beta = d$. Based on the distribution theory of Section 2 and the algebraic simplifications of Section 3, the $(1-\alpha)100\%$ confidence region consists of all the vectors d that satisfy the inequality

$$\frac{[\Lambda'\hat{\beta} - d]'[\Lambda'(X'X)^-\Lambda]^-[\Lambda'\hat{\beta} - d]/r(\Lambda)}{MSE} \leq F(1-\alpha, r(\Lambda), r(I-M)). \qquad (1)$$

These vectors form an ellipsoid in $r(\Lambda)$-dimensional space.

Alternative forms for the confidence region are

$$[\Lambda'\hat{\beta} - d]'[\Lambda'(X'X)^-\Lambda]^-[\Lambda'\hat{\beta} - d] \leq r(\Lambda)\, MSE\, F(1-\alpha, r(\Lambda), r(I-M))$$

and

$$[P'MY - d]'(P'MP)^-[P'MY - d] \leq r(MP)\, MSE\, F(1-\alpha, r(MP), r(I-M)).$$

For regression problems we can get a considerable simplification. If we take $P' = (X'X)^{-1}X'$, then we have $\Lambda' = P'X = I_p$ and $\Lambda'\beta = \beta = d$. Using these in (1) and renaming the placeholder variable d as β gives

$$\begin{aligned}
[\Lambda'\hat{\beta} - d]'[\Lambda'(X'X)^-\Lambda]^-[\Lambda'\hat{\beta} - d] &= (\hat{\beta} - \beta)'[(X'X)^{-1}]^{-1}(\hat{\beta} - \beta) \\
&= (\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta)
\end{aligned}$$

with $r(\Lambda) = r(I_p) = r(X)$. The confidence region is thus the set of all βs satisfying

$$(\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta) \leq p\, MSE\, F(1-\alpha, p, n-p).$$

3.8 Tests for Generalized Least Squares Models

We now consider the problem of deriving tests for the model of Section 2.7. For testing, we take the generalized least squares model as

$$Y = X\beta + e, \quad e \sim N\left(0, \sigma^2 V\right), \tag{1}$$

where V is a known positive definite matrix. As in Section 2.7, we can write $V = QQ'$ for Q nonsingular. The model

$$Q^{-1}Y = Q^{-1}X\beta + Q^{-1}e, \quad Q^{-1}e \sim N\left(0, \sigma^2 I\right) \tag{2}$$

is analyzed instead of model (1).

First consider the problem of testing model (1) against a reduced model, say

$$Y = X_0\beta_0 + e, \quad e \sim N\left(0, \sigma^2 V\right), \quad C(X_0) \subset C(X). \tag{3}$$

The reduced model can be transformed to

$$Q^{-1}Y = Q^{-1}X_0\beta_0 + Q^{-1}e, \quad Q^{-1}e \sim N\left(0, \sigma^2 I\right). \tag{4}$$

The test of model (3) against model (1) is performed by testing model (4) against model (2). To test model (4) against model (2), we need to know that model (4) is a reduced model relative to model (2). In other words, we need to show that $C(Q^{-1}X_0) \subset C(Q^{-1}X)$. From model (3), $C(X_0) \subset C(X)$, so there exists a matrix U such that $X_0 = XU$. It follows immediately that $Q^{-1}X_0 = Q^{-1}XU$; hence $C(Q^{-1}X_0) \subset C(Q^{-1}X)$.

Recall from Section 2.7 that $A = X(X'V^{-1}X)^{-}X'V^{-1}$ and that for model (1)

$$MSE = Y'(I-A)'V^{-1}(I-A)Y \Big/ [n - r(X)].$$

Define $A_0 = X_0(X_0'V^{-1}X_0)^{-}X_0'V^{-1}$. The test comes from the following distributional result.

Theorem 3.8.1.

(i) $$\frac{Y'(A-A_0)'V^{-1}(A-A_0)Y/[r(X)-r(X_0)]}{MSE} \sim F(r(X)-r(X_0), n-r(X), \pi),$$

where $\pi = \beta'X'(A-A_0)'V^{-1}(A-A_0)X\beta/2\sigma^2$.

(ii) $\beta'X'(A-A_0)'V^{-1}(A-A_0)X\beta = 0$ if and only if $\mathrm{E}(Y) \in C(X_0)$.

PROOF.

(i) Theorem 3.2.1 applied to models (2) and (4) gives the appropriate test statistic. It remains to show that part (i) involves the same test statistic. Exercise 3.8 is to show that $Y'(A-A_0)'V^{-1}(A-A_0)Y/[r(X)-r(X_0)]$ is the appropriate numerator mean square.

(ii) From part (i) and Theorem 3.2.1 applied to models (2) and (4),

$$\beta'X'(A-A_0)'V^{-1}(A-A_0)X\beta = 0$$

if and only if $\mathrm{E}(Q^{-1}Y) \in C(Q^{-1}X_0)$. $\mathrm{E}(Q^{-1}Y) \in C(Q^{-1}X_0)$ if and only if $\mathrm{E}(Y) \in C(X_0)$. □

Exercise 3.8 Show that $Y'(A-A_0)'V^{-1}(A-A_0)Y/[r(X)-r(X_0)]$ is the appropriate numerator mean square for testing model (4) against model (2).

The intuition behind the test based on Theorem 3.8.1 is essentially the same as that behind the usual test (which was discussed in Section 2). The usual test is based on the difference $MY - M_0Y = (M-M_0)Y$. MY is the estimate of $\mathrm{E}(Y)$ from the full model, and M_0Y is the estimate of $\mathrm{E}(Y)$ from the reduced model. The difference between these estimates indicates how well the reduced model fits. If the difference is large, the reduced model fits poorly; if the difference is small, the reduced model fits relatively well. To determine whether the difference vector is large or small, the squared length of the vector, as measured in Euclidean distance, is used. The squared length of $(M-M_0)Y$ reduces to the usual form $Y'(M-M_0)Y$. The basis of the test is to quantify how large the difference vector must be before there is some assurance that the difference between the vectors is due to more than just the variability of the data.

For generalized least squares models, the estimate of $\mathrm{E}(Y)$ from the full model is AY and the estimate of $\mathrm{E}(Y)$ from the reduced model is A_0Y. The difference between these vectors, $AY - A_0Y = (A-A_0)Y$, indicates how well the reduced model fits. The test is based on the squared length of the vector $(A-A_0)Y$, but the length of the vector is no longer measured in terms of Euclidean distance. The inverse of the covariance matrix is used to define a distance measure appropriate to generalized least squares models. Specifically, the squared distance between two vectors u and v is defined to be $(u-v)'V^{-1}(u-v)$. Note that with this distance measure, the generalized least squares estimate AY is the vector in $C(X)$ that is closest to Y, i.e., AY is the perpendicular projection onto $C(X)$ (cf. Section 2.7).

It should be noted that if $V = I$, then $A = M$, $A_0 = M_0$, and the test is exactly as in Section 2. Also as in Section 2, the key term in the numerator of the test statistic, $Y'(A-A_0)'V^{-1}(A-A_0)Y$, can be obtained as the difference between the SSE for the reduced model and the SSE for the full model.

We now consider testing parametric functions. If $\Lambda'\beta$ is an estimable parametric vector, then the test of the hypothesis $\Lambda'\beta = 0$ can be obtained from the following result:

Theorem 3.8.2.

(i) $$\frac{\hat{\beta}'\Lambda\left[\Lambda'(X'V^{-1}X)^{-}\Lambda\right]^{-}\Lambda'\hat{\beta}/r(\Lambda)}{MSE} \sim F(r(\Lambda), n-r(X), \pi),$$

where $\pi = \beta'\Lambda\left[\Lambda'(X'V^{-1}X)^{-}\Lambda\right]^{-}\Lambda'\beta/2\sigma^2$.

(ii) $\beta'\Lambda\left[\Lambda'(X'V^{-1}X)^{-}\Lambda\right]^{-}\Lambda'\beta = 0$ if and only if $\Lambda'\beta = 0$.

PROOF. $\Lambda'\beta$ is estimable in model (1) if and only if $\Lambda'\beta$ is estimable in model (2). $\Lambda'\hat{\beta}$ is the least squares estimate of $\Lambda'\beta$ from model (2), and $\sigma^2\Lambda'(X'V^{-1}X)^{-}\Lambda$ is the covariance matrix of $\Lambda'\hat{\beta}$. The result follows immediately from Section 3 applied to model (2). □

Note that $\Lambda'\beta = 0$ defines the same reduced model as in Section 3 but the test of the reduced model changes. Just as in Section 3 for ordinary least squares models, Theorem 3.8.2 provides a method of finding tests for generalized least squares models. To test $\Lambda'\beta = 0$, one need only find $\Lambda'\hat{\beta}$, $\mathrm{Cov}(\Lambda'\hat{\beta})$, and MSE. If these can be found, the test follows immediately.

We have assumed that V is a known matrix. Since the results depend on V, they would seem to be of little use if V were not known. Nevertheless, the validity of the results does not depend on V being known. In Chapter 11, we will consider cases where V is not known, but where V and X are related in such a way that the results of this section can be used. In Chapter 11, we will need the distribution of the numerators of the test statistics.

Theorem 3.8.3.

(i) $Y'(A-A_0)'V^{-1}(A-A_0)Y/\sigma^2 \sim \chi^2(r(X)-r(X_0),\pi)$,

where $\pi = \beta'X'(A-A_0)'V^{-1}(A-A_0)X\beta/2\sigma^2$, and $\beta'X'(A-A_0)'V^{-1}(A-A_0)X\beta = 0$ if and only if $\mathrm{E}(Y) \in C(X_0)$.

(ii) $\hat{\beta}'\Lambda\left[\Lambda'(X'V^{-1}X)^{-}\Lambda\right]^{-}\Lambda'\hat{\beta}/\sigma^2 \sim \chi^2(r(\Lambda),\pi)$,

where $\pi = \beta'\Lambda\left[\Lambda'(X'V^{-1}X)^{-}\Lambda\right]^{-}\Lambda'\beta/2\sigma^2$, and $\Lambda'\beta = 0$ if and only if $\pi = 0$.

PROOF. The results follow from Sections 3.2 and 3.3 applied to model (2). □

Exercise 3.9 Show that $Y'(A-A_0)'V^{-1}(A-A_0)Y$ equals the difference in the SSEs for models (3) and (1).

3.8.1 Conditions for Simpler Procedures

Just as Proposition 2.7.5 establishes that least squares estimates can be BLUEs even when $\mathrm{Cov}(Y) \equiv \sigma^2 V \neq \sigma^2 I$, there exist conditions where $\mathrm{Cov}(Y) \equiv \sigma^2 V \neq \sigma^2 I$ but under which the F statistic of Section 3.2 still has an $F(r(M-M_0), r(I-M))$ distribution under the null hypothesis with multivariate normal errors. In particular, when testing $Y = X\beta + e$ versus the reduced model $Y = X_0\gamma + e$ where $\mathrm{Cov}(Y) \equiv$

$V = \sigma^2[I + X_0 B' + BX_0']$, the standard central F distribution continues to hold under the null hypothesis. In fact, the matrix B can even contain unknown parameters without affecting the validity of this result.

To see the result on F tests, one need only check that the usual numerator and denominator have the same independent χ^2 distributions under the null hypothesis as established in the proof of Theorem 3.2.1. This can be demonstrated by applying Theorems 1.3.6 and 1.3.8. In particular, $Y'(M - M_0)Y/\sigma^2 \sim \chi^2(r(M - M_0))$ because, with this V,

$$\left[\frac{1}{\sigma^2}(M - M_0)\right] V \left[\frac{1}{\sigma^2}(M - M_0)\right] = \frac{1}{\sigma^2}(M - M_0).$$

Similarly, $Y'(I - M)Y/\sigma^2 \sim \chi^2(r(I - M))$ because

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

Finally, independence follows from the fact that

$$\left[\frac{1}{\sigma^2}(M - M_0)\right] V \left[\frac{1}{\sigma^2}(I - M)\right] = 0.$$

Moreover, the arguments given in Huynh and Feldt (1970) should generalize to establish that the F distribution holds only if V has the form indicated.

Of course, it is not clear whether $\sigma^2[I + X_0 B' + BX_0']$ is positive definite, as good covariance matrices should be. However, $X_0 B' + BX_0'$ is symmetric, so it has real eigenvalues, and if the negative of its smallest eigenvalue is less than 1, $I + X_0 B' + BX_0'$ will be positive definite. A special case of this covariance structure has $X_0 B' + BX_0' = X_0 B_0 X_0'$ for some B_0. In this special case, it is enough to have B_0 nonnegative definite. Also in this special case, not only do standard F tests apply, but least squares estimates are BLUEs because $C(VX) \subset C(X)$, so Proposition 2.7.5 applies. But in general with $V = I + X_0 B' + BX_0'$, it is possible to use the standard F tests even though least squares does not give BLUEs.

To illustrate the ideas, consider a balanced two-way ANOVA without interaction or replication, $y_{ij} = \mu + \alpha_i + \eta_j + e_{ij}$, $i = 1, \ldots, a$, $j = 1, \ldots, b$. *In this context, we think about the α_is as block effects, so there are a blocks and b treatments.* We explore situations in which observations within each block are correlated, but the usual F test for treatment effects continues to apply. Write the linear model in matrix form as

$$Y = X\beta + e = [J_{ab}, X_\alpha, X_\eta] \begin{bmatrix} \mu \\ \alpha \\ \eta \end{bmatrix} + e.$$

Here $\alpha = (\alpha_1, \ldots, \alpha_a)'$ and $\eta = (\eta_1, \ldots, \eta_b)'$ The test of no treatment effects uses the reduced model

$$Y = X_0\gamma + e = [J_{ab}, X_\alpha] \begin{bmatrix} \mu \\ \alpha \end{bmatrix} + e.$$

In the first illustration given below, $V = I + X_\alpha B_{0*} X'_\alpha$ for some B_{0*}. In the second illustration, $V = I + X_\alpha B'_* + B_* X'_\alpha$. In both cases it suffices to write V using X_α rather than X_0. This follows because $C(X_0) = C(X_\alpha)$, so we can always write $X_0 B' = X_\alpha B'_*$.

One covariance structure that is commonly used involves *compound symmetry*, that is, independence between blocks, homoscedasticity, and constant correlation within blocks. In other words,

$$\mathrm{Cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_*^2 & \text{if } i = i', j = j' \\ \sigma_*^2 \rho & \text{if } i = i', j \neq j' \\ 0 & \text{if } i \neq i'\ . \end{cases}$$

One way to write this covariance matrix is as

$$\sigma^2 V = \sigma_*^2 (1 - \rho) I + \sigma_*^2 \rho X_\alpha X'_\alpha .$$

In the context, σ^2 from the general theory is $\sigma_*^2(1-\rho)$ from the example and $B_{0*} \equiv [\rho/(1-\rho)] I_a$.

A more general covariance structure is

$$\mathrm{Cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma^2(1 + 2\delta_j) & \text{if } i = i', j = j' \\ \sigma^2(\delta_j + \delta_{j'}) & \text{if } i = i', j \neq j' \\ 0 & \text{if } i \neq i'\ . \end{cases}$$

We want to find B so that this covariance structure can be written as $\sigma^2[I + X_0 B' + B X'_0]$. It suffices to show that for some B_* we can write $V = I + X_\alpha B'_* + B_* X'_\alpha$. In the balanced two-way ANOVA without interaction or replication, when $Y = [y_{11}, y_{12}, \ldots, y_{ab}]'$, X_α can be written using Kronecker products as

$$X_\alpha = [I_a \otimes J_b] = \begin{bmatrix} J_b & 0 & \cdots & 0 \\ 0 & J_b & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & J_b \end{bmatrix} .$$

Now define $\delta = (\delta_1, \ldots, \delta_b)'$ and take

$$B_* = [I_a \otimes \delta] = \begin{bmatrix} \delta & 0 & \cdots & 0 \\ 0 & \delta & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \delta \end{bmatrix} .$$

With these choices, it is not difficult to see that the covariance structure specified earlier has

$$V = I_{ab} + [I_a \otimes J_b][I_a \otimes \delta]' + [I_a \otimes \delta][I_a \otimes J_b]' .$$

This second illustration is similar to a discussion in Huynh and Feldt (1970). The split plot models of Chapter 11 involve covariance matrices with compound symmetry, so they are *similar* in form to the first illustration that involved $\sigma^2 I + X_0 B_0 X_0'$. The results here establish that the F tests in the subplot analyses of Chapter 11 could still be obtained when using the more general covariance structures considered here.

3.9 Additional Exercises

Exercise 3.9.1 Consider the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$, e_is i.i.d. $N(0, \sigma^2)$. Use the data given below to answer (a) and (b).

Obs.	1	2	3	4	5	6
y	-2	7	2	5	8	-1
x_1	4	-1	2	0	-2	3
x_2	2	-3	0	-2	-4	1

(a) Find $SSR(X_1, X_2 | J) = R(\beta_1, \beta_2 | \beta_0)$.
(b) Are β_0, β_1, and β_2 estimable?

Exercise 3.9.2 For a standard linear model, find the form of the generalized likelihood ratio test of $H_0 : \sigma^2 = \sigma_0^2$ versus $H_A : \sigma^2 \neq \sigma_0^2$ in terms of rejecting H_0 when some function of SSE/σ_0^2 is small. Show that the test makes sense in that it rejects for both large and small values of SSE/σ_0^2.

Exercise 3.9.3 Consider a set of seemingly unrelated regression equations

$$Y_i = X_i \beta_i + e_i, \quad e_i \sim N\left(0, \sigma^2 I\right),$$

$i = 1, \ldots, r$, where X_i is an $n_i \times p$ matrix and the e_is are independent. Find the test for $H_0 : \beta_1 = \cdots = \beta_r$.

Exercise 3.9.4 What happens to the test of $\Lambda'\beta = d$ if $\Lambda'\beta$ has no estimable part?

Exercise 3.9.5 Consider the model

$$Y = X\beta + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 I, \tag{1}$$

with the additional restriction

$$\Lambda'\beta = d,$$

where $d = \Lambda' b$ for some (known) vector b and $\Lambda' = P'X$. Model (1) with the additional restriction is equivalent to the model

$$(Y - Xb) = (M - M_{MP})\gamma + e. \tag{2}$$

If the parameterization of model (1) is particularly appropriate, then we might be interested in estimating $X\beta$ subject to the restriction $\Lambda'\beta = d$. To do this, write

$$X\beta = \mathrm{E}(Y) = (M - M_{MP})\gamma + Xb,$$

and define the BLUE of $\lambda'\beta = \rho' X\beta$ in the restricted version of (1) to be $\rho'(M - M_{MP})\hat{\gamma} + \rho' Xb$, where $\rho'(M - M_{MP})\hat{\gamma}$ is the BLUE of $\rho'(M - M_{MP})\gamma$ in model (2). Let $\hat{\beta}_1$ be the least squares estimate of β in the unrestricted version of model (1). Show that the BLUE of $\lambda'\beta$ in the restricted version of model (1) is

$$\lambda'\hat{\beta}_1 - \left[\mathrm{Cov}(\lambda'\hat{\beta}_1, \Lambda'\hat{\beta}_1)\right]\left[\mathrm{Cov}(\Lambda'\hat{\beta}_1)\right]^{-}(\Lambda'\hat{\beta}_1 - d), \tag{3}$$

where the covariance matrices are computed as in the unrestricted version of model (1).

Hint: This exercise is actually nothing more than simplifying the terms in (3) to show that it equals $\rho'(M - M_{MP})\hat{\gamma} + \rho' Xb$.

Note: The result in (3) is closely related to best linear prediction, cf. Sections 6.3 and 12.2.

Exercise 3.9.6 Discuss how the testing procedures from this chapter would change if you actually knew the variance σ^2.