Chapter 11
Split Plot Models

In an experiment with at least two factors, it is sometimes convenient to apply some of the factors to large experimental units (called *whole plots*) and then to split the large units into smaller parts to which the remaining factors are applied. The subdivisions of the whole plots are called *subplots* or *split plots*.

Split plot designs are often used when either (1) the factors applied to whole plots are not of direct interest or (2) some factors require larger experimental units than the other factors. The first case is illustrated with an experiment to evaluate crop yields when using varying levels of a standard herbicide and a new pesticide. If the standard herbicide is not of direct interest, but rather primary interest is in the effects of the pesticide and any possible interaction between the herbicide and the pesticide, then it is appropriate to apply herbicides as whole plot treatments. (It will be seen later that interaction contrasts and comparisons between pesticides are subject to less error than comparisons among herbicides.) The second case that split plot designs are often used for can also be illustrated with this experiment. If the standard herbicide is applied using a tractor, but the new pesticide is applied by crop dusting, then the experimental procedure makes it necessary to use pesticides as whole plot treatments. (Clearly, an airplane requires a larger plot of ground for spraying than does a tractor.)

It is of interest to note that a split plot design can be thought of as an (unbalanced) incomplete block design. In this approach, each whole plot is thought of as a block. Each block contains the treatments that are all combinations of the subplot factor levels with the one combination of whole plot factor levels that was applied. As with other incomplete block designs, a split plot design is necessary when there are not enough blocks available that can accommodate all of the treatment combinations. In split plot designs, this means that there are not enough subplots per whole plot so that all treatment combinations could be applied at the subplot level. If, in addition, there are not enough whole plots so that each treatment combination could be applied to a whole plot, then a split plot design is an attractive option.

Mathematically, the key characteristic of a split plot model is the covariance structure. Typically, observations taken on the subplots of any particular whole plot

R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*,
Springer Texts in Statistics, DOI 10.1007/978-1-4419-9816-3_11,

are assumed to have a constant nonzero correlation. Observations taken on different whole plots are assumed to be uncorrelated.

The main purpose of this chapter is to derive the analysis for split plot models. In Section 1, we consider a special cluster sampling model. The cluster sampling model has the same covariance structure as a split plot model. In Section 2, we consider ways of generalizing the cluster sampling model that allow for an easy analysis of the data. The discussion in Section 2 is really an examination of generalized split plot models. Section 3 derives the analysis for the traditional split plot model by using the results of Section 2. Section 4 discusses the issues of identifying an appropriate error term and of subsampling.

Sections 1 and 2 are closely related to Christensen (1984) and (1987b), respectively. In fact, Christensen (1987b) is probably easier to read than Section 2 because it includes more introductory material and fewer of the mathematical details. Closely related work is contained in Monlezun and Blouin (1988) and Mathew and Sinha (1992). A general review of methods for analyzing cluster sampling models is given in Skinner, Holt, and Smith (1989).

11.1 A Cluster Sampling Model

A commonly used technique in survey sampling is *cluster sampling* (also called *two-stage sampling*). This technique is applied when the population to be sampled consists of some kind of clusters. The sample is obtained by taking a random sample of clusters and then taking a random sample of the individuals within each of the sampled clusters. For example, suppose it was desired to sample the population of grade school students in Montana. One could take a random sample of grade schools in the state, and then for each school that was chosen take a random sample of the students in the school. One complication of this method is that students from the same school will tend to be more alike than students from different schools. In general, there will be a nonnegative correlation among the individual units within a cluster.

General Cluster Sampling Models

Suppose n observations are available from a two-stage sample with c clusters. From each cluster, m_i units are sampled and variables $Y, X_1, \ldots, X_p$ are obtained. Since observations in a cluster are typically not independent, we will consider the linear model

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

where X is $n \times p$ of rank r and (assuming the elements of Y are listed by clusters) V is the block diagonal matrix

$$V = \text{Blk diag}(V_i),$$

where V_i is an $m_i \times m_i$ *intraclass correlation* matrix

$$V_i = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}.$$

If we let $J_{m(i)}$ be an $m_i \times 1$ vector of ones, then

$$V = (1-\rho)I + \rho\,\text{Blk diag}(J_{m(i)}J'_{m(i)}).$$

Now let X_1 be an $n \times c$ matrix of indicator variables for the clusters. In other words, a row of the ith column of X_1 is 1 if the row corresponds to an observation from the ith cluster and 0 otherwise. It follows that $X_1X_1' = \text{Blk diag}(J_{m(i)}J'_{m(i)})$, so

$$V = (1-\rho)I + \rho X_1X_1'. \tag{1}$$

In fact, equation (1) holds even if the elements of Y are not listed by cluster.

We can now provide an interesting condition for when ordinary least squares (OLS) estimates are best linear unbiased estimates (BLUEs) in cluster sampling models. Recall from Theorem 10.4.5 that OLS estimates are BLUEs if and only if $C(VX) \subset C(X)$. This condition holds if and only if $C(X_1X_1'X) \subset C(X)$. Since the columns of X_1 are indicator variables for the clusters, $X_1X_1'X$ takes each column of X, computes the cluster totals, and replaces each component with the corresponding cluster total. Thus, OLS estimates are BLUEs if and only if for any variable in the model, the variable formed by replacing each component with the corresponding cluster total is also, either implicitly or explicitly, contained in the model.

A Special Cluster Sampling Model

We now consider a particular cluster sampling model for which OLS estimates are BLUEs, and for which tests and confidence intervals are readily available for the most interesting parameters. Consider a model in which X can be written as $X = [X_1, X_2]$, where X_1 is again the matrix of indicator variables for the clusters. Rewriting the linear model as

$$Y = [X_1, X_2]\begin{bmatrix} \alpha \\ \gamma \end{bmatrix} + e \tag{2}$$

leads to the interpretation that the α_is are separate cluster effects. Typically, one would not be very interested in these cluster effects. One's primary interest would be in the vector γ.

It is easily seen that $C(VX) \subset C(X)$, so OLS estimates are BLUEs. The noteworthy thing about this model is that inference on the parameter vector γ can proceed just as when $\text{Cov}(Y) = \sigma^2 I$. Treating (2) as an analysis of covariance model, we obtain for any estimable function $\lambda'\gamma$

$$\lambda'\hat{\gamma} = \lambda' \left[X_2'(I-M_1)X_2\right]^- X_2'(I-M_1)Y,$$

where M_1 is the perpendicular projection matrix onto $C(X_1)$.

The variance of $\lambda'\hat{\gamma}$ is

$$\text{Var}(\lambda'\hat{\gamma}) = \sigma^2\lambda' \left[X_2'(I-M_1)X_2\right]^- X_2'(I-M_1)V(I-M_1)X_2 \left[X_2'(I-M_1)X_2\right]^- \lambda. \quad (3)$$

From (1) observe that

$$V(I-M_1) = (1-\rho)(I-M_1).$$

Substitution into (3) gives

$$\text{Var}(\lambda'\hat{\gamma}) = \sigma^2(1-\rho)\lambda' \left[X_2'(I-M_1)X_2\right]^- \lambda, \quad (4)$$

which, except for the term $(1-\rho)$, is the variance from assuming $\text{Cov}(Y) = \sigma^2 I$.

Exercise 11.1 Prove that equation (4) is true.

The mean square error (MSE) from ordinary least squares provides an independent estimate of $\sigma^2(1-\rho)$. Let $M = X(X'X)^-X'$, so $MSE = Y'(I-M)Y/(n-r)$.

$$\text{E}(MSE) = (n-r)^{-1}\sigma^2 \text{tr}[(I-M)V].$$

Since $C(X_1) \subset C(X)$, from (1) we have

$$(I-M)V = (I-M)(1-\rho)I = (1-\rho)(I-M).$$

But, $\text{tr}[(1-\rho)(I-M)] = (1-\rho)(n-r)$, so

$$\text{E}(MSE) = \sigma^2(1-\rho).$$

Theorem 11.1.1.

(i) $Y'(I-M)Y/\sigma^2(1-\rho) \sim \chi^2(n-r,0)$.
(ii) MSE and $X\hat{\beta}$ are independent. In particular, MSE and $\lambda'\hat{\gamma}$ are independent.

Exercise 11.2 Prove Theorem 11.1.1.
Hint: For (i), use Theorem 1.3.6. For (ii), show that $\text{Cov}[(I-M)Y, MY] = 0$.

These results provide a basis for finding tests and confidence intervals for an estimable function $\lambda'\gamma$. We might also want to consider doing F tests. Suppose we want to test some vector of estimable restrictions on γ, say $\Lambda'\gamma = 0$. The test can be derived from $\Lambda'\hat{\gamma}$, $\mathrm{Cov}(\Lambda'\hat{\gamma})$, and MSE, using Theorem 11.1.1 and Corollary 3.8.3. In particular,

Theorem 11.1.2.

(i) $$\frac{(\Lambda'\hat{\gamma})'\left(\Lambda'[X_2'(I-M_1)X_2]^-\Lambda\right)^-(\Lambda'\hat{\gamma})/r(\Lambda)}{MSE} \sim F(r(\Lambda), n-r, \pi),$$

where $\pi = (\Lambda'\gamma)'\left(\Lambda'[X_2'(I-M_1)X_2]^-\Lambda\right)^-(\Lambda'\gamma)/2\sigma^2(1-\rho)$.

(ii) $(\Lambda'\gamma)'\left(\Lambda'[X_2'(I-M_1)X_2]^-\Lambda\right)^-(\Lambda'\gamma) = 0$ if and only if $\Lambda'\gamma = 0$.

An alternative to testing linear parametric functions is testing models. To test model (2) against a reduced model, say

$$Y = X_0\beta_0 + e, \quad C(X_1) \subset C(X_0) \subset C(X),$$

the test is the usual $\mathrm{Cov}(Y) = \sigma^2 I$ test. With $M_0 = X_0(X_0'X_0)^-X_0'$, we have:

Theorem 11.1.3.

(i) $$\frac{Y'(M-M_0)Y/[r(X)-r(X_0)]}{MSE} \sim F(r(X)-r(X_0), n-r, \pi),$$

where $\pi = \beta_0'X'(M-M_0)X\beta_0/2\sigma^2(1-\rho)$.

(ii) $\beta_0'X'(M-M_0)X\beta_0 = 0$ if and only if $\mathrm{E}(Y) \in C(X_0)$.

PROOF. For part (i), see Exercise 11.3. Part (ii) follows exactly as in Theorem 3.2.1. □

Exercise 11.3 Prove Theorem 11.1.3(i).

In summary, for a model that includes separate fixed effects for each cluster, the ordinary least squares fit gives optimal estimates of all effects and valid estimates of standard errors for all effects not involving the cluster effects. If normal distributions are assumed, the usual $\mathrm{Cov}(Y) = \sigma^2 I$ tests and confidence intervals are valid unless the cluster effects are involved. If the cluster effects are not of interest, the entire analysis can be performed with ordinary least squares. This substantially reduces the effort required to analyze the data.

The assumption that the α_is are fixed effects is necessary for the result to hold. However, if additional random cluster effects are added to the model so that there

are both fixed and random cluster effects, then the basic structure of the covariance matrix remains unchanged and the optimality of ordinary least squares is retained.

Exercise 11.4 The usual model for a randomized complete block design was given in Section 8.2 as

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

$i = 1, \ldots, a$, $j = 1, \ldots, b$, $\text{Var}(e_{ij}) = \sigma^2$, and $\text{Cov}(e_{ij}, e_{i'j'}) = 0$ for $(i, j) \neq (i', j')$. The β_js are considered as fixed block effects. Consider now a model

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

The η_js are independent $N(0, \sigma_2^2)$ and the e_{ij}s are independent $N(0, \sigma_1^2)$. The η_js and e_{ij}s are also independent. The block effects are now $(\beta_j + \eta_j)$. There is a fixed component and a random component with mean zero in each block effect. Use the results of this section to derive an analysis for this model. Give an ANOVA table and discuss interval estimates for contrasts in the α_is.

Exercise 11.5 An alternative model for a block design is

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \tag{5}$$

where the β_js are independent $N(0, \sigma_2^2)$ and the β_js and e_{ij}s are independent. If this model is used for a balanced incomplete block design, the BLUEs are different than they are when the β_js are assumed to be fixed. Discuss the appropriateness of the analysis based on this model in light of the results of Exercise 11.4.

Exercise 11.6 Show that when using model (5) for a randomized complete block design, the BLUE of a contrast in the α_is is the same regardless of whether the β_js are assumed random or fixed. Show that the estimates of a contrast's variance are the same.

11.2 Generalized Split Plot Models

By placing additional conditions on model (11.1.2), we can get a simple analysis of the cluster effects while retaining a simple analysis for the noncluster effects. The analysis of the cluster effects corresponds to the analysis of whole plot treatments in a split plot model. The noncluster effects relate to effects on the subplot level.

Generalized split plot models are models obtained by imposing additional structure on the cluster sampling model (11.1.2). This additional structure involves simplifying the covariance matrix and modeling the whole plot (cluster) effects. First, a condition on the whole plots is discussed. The condition is that the number of observations in each whole plot is the same. This condition simplifies the covari-

ance matrix considerably. Next, the whole plot effects are modeled by assuming a reduced model that does not allow separate effects for each whole plot. As part of the modeling process, a condition is imposed on the model matrix of the reduced model that ensures that least squares estimates are BLUEs. The problem of drawing inferences about generalized split plot models is discussed in two parts, (1) estimation and testing of estimable functions and (2) testing reduced models. A condition that allows for a simple analysis of the whole plot effects is mentioned, a discussion of how to identify generalized split plot models is given, and finally some computational methods are presented.

The Covariance Matrix

Writing Y so that observations in each whole plot are listed contiguously, we can rewrite the covariance matrix of model (11.1.2) as

$$\sigma^2 V = \sigma^2 \left[(1-\rho)I + \rho M_1 \{ \text{Blk diag}(m_i I_{m(i)}) \} \right],$$

where $I_{m(i)}$ is an $m_i \times m_i$ identity matrix and M_1 is the perpendicular projection matrix onto $C(X_1)$. This follows from Section 1 because $M_1 = \text{Blk diag}(m_i^{-1} J_{m(i)} J'_{m(i)})$. This characterization of V is not convenient in itself because of the Blk diag$(m_i I_{m(i)})$ term. For example, the expected value of a quadratic form, say $Y'AY$, is $\mathrm{E}(Y'AY) = \sigma^2 \text{tr}(AV) + \beta' X' A X \beta$. The trace of AV is not easy to compute. To simplify the subsequent analysis, we impose

Condition 11.2.1. All whole plots (clusters) are of the same size, say $m_i = m$ for all i.

It follows that

$$V = (1-\rho)I + m\rho M_1. \tag{1}$$

This form for V will be assumed in the remainder of Section 2.

Modeling the Whole Plots

Model (11.1.2) assumes separate effects for each whole plot (cluster). Modeling the cluster effects consists of imposing structure on those effects. This is done by putting a constraint on $C(X_1)$. The simplest way to do this is by postulating a reduced model, say

$$Y = Z\beta_* + e, \quad Z = [X_*, X_2], \quad C(X_*) \subset C(X_1). \tag{2}$$

Partitioning β_* in conformance with Z, write

$$\beta_*' = [\delta', \gamma'].$$

Remember, this is not the same γ as in (11.1.2), but it is the coefficient for X_2, just as in (11.1.2). Define the perpendicular projection operators onto $C(Z)$ and $C(X_*)$ as M_Z and M_*, respectively.

In Section 1, it was shown that least squares estimates were BLUEs for model (11.1.2). We are now dealing with a different model, model (2), so another proof is required. To check if least squares estimates are BLUEs, we need to see whether $C(VZ) \subset C(Z)$. Since by equation (1), $V = (1-\rho)I + m\rho M_1$, we have $VZ = (1-\rho)Z + m\rho M_1 Z$. Clearly, it is enough to check whether $C(M_1 Z) \subset C(Z)$. This is true for a special case.

Proposition 11.2.2. Let $\mathscr{M}$ and $\mathscr{N}$ be subspaces of $C(Z)$. If $C(Z) = \mathscr{M} + \mathscr{N}$, where $\mathscr{M} \subset C(X_1)$ and $\mathscr{N} \perp C(X_1)$, then $C(M_1 Z) \subset C(Z)$.

PROOF. For any $v \in C(Z)$, write $v = v_1 + v_2$, where $v_1 \in \mathscr{M}$ and $v_2 \in \mathscr{N}$. $M_1 v = M_1 v_1 + M_1 v_2 = v_1$, but $v_1 \in \mathscr{M} \subset C(Z)$. □

A condition that is easy to check is

Condition 11.2.3. $C(Z) = C(X_*, (I - M_1)X_2)$ and $C(X_*) \subset C(X_1)$.

If Condition 11.2.3 holds, then Proposition 11.2.2 applies with $\mathscr{M} = C(X_*)$ and $\mathscr{N} = C[(I - M_1)X_2]$. If Proposition 11.2.2 applies, then least squares estimates are BLUEs.

EXAMPLE 11.2.4. Let whole plots be denoted by the subscripts i and j, and let subplots have the subscript k. Let the dependent variable be y_{ijk} and let x_{ijk1}, x_{ijk2}, and x_{ijk3} be three covariates. The model given below is a generalized split plot model (see Exercise 11.7.):

$$\begin{aligned} y_{ijk} = {} & \mu + \omega_i + \gamma_1 \bar{x}_{ij\cdot 1} + \gamma_{21} \bar{x}_{ij\cdot 2} + \eta_{ij} \\ & + \tau_k + (\omega\tau)_{ik} + \gamma_{22}(x_{ijk2} - \bar{x}_{ij\cdot 2}) + \gamma_3(x_{ijk3} - \bar{x}_{ij\cdot 3}) \\ & + e_{ijk}, \end{aligned} \tag{3}$$

$i = 1, \ldots, a$, $j = 1, \ldots, N_i$, $k = 1, \ldots, m$. The η_{ij}s and e_{ijk}s are all independent with $\eta_{ij} \sim N(0, \sigma_w^2)$ and $e_{ijk} \sim N(0, \sigma_s^2)$. With these assumptions

$$\sigma^2 = \sigma_w^2 + \sigma_s^2$$

and

$$\rho = \sigma_w^2 / (\sigma_w^2 + \sigma_s^2).$$

The ω_is are treatment effects for a one-way ANOVA with unequal numbers in the whole plots. The whole plot treatments can obviously be generalized to include

multifactor ANOVAs with unequal numbers. The ω_is, γ_1, γ_{21}, and μ make up the δ vector. The τ_ks, $(\omega\tau)_{ik}$s, γ_{22}, and γ_3 make up the vector γ from model (2). Note that the covariate used with γ_{22} could be changed to x_{ijk2} without changing $C(Z)$ or invalidating Condition 11.2.3.

Exercise 11.7 Verify that Condition 11.2.3 holds for model (3).

Estimation and Testing of Estimable Functions

We now discuss estimation and testing for model (2). Under Condition 11.2.1 and Condition 11.2.3, least squares estimates are BLUEs. Define

$$M_2 = (I - M_1)X_2 \left[X_2'(I - M_1)X_2\right]^- X_2'(I - M_1).$$

From Condition 11.2.3, the perpendicular projection operator onto $C(Z)$ is

$$M_Z = M_* + M_2. \tag{4}$$

Given the perpendicular projection operator, the least squares estimates can be found in the usual way.

First, consider drawing inferences about γ. For estimable functions of γ, the estimates are exactly as in Section 1. In both cases, the estimates depend only on M_2Y. (See Proposition 9.1.1.) Since model (11.1.2) is a larger model than model (2) [i.e., $C(Z) \subset C(X)$], model (11.1.2) remains valid. It follows that all of the distributional results in Section 1 remain valid. In particular,

$$Y'(I - M)Y/\sigma^2(1-\rho) \sim \chi^2(n - r(X), 0), \tag{5}$$

and $Y'(I-M)Y/[n-r(X)]$ is an unbiased estimate of $\sigma^2(1-\rho)$. In split plot models, $\sigma^2(1-\rho)$ is called the *subplot error variance*. $Y'(I-M)Y$ is called the *sum of squares for subplot error* $[SSE(s)]$ and $Y'(I-M)Y/[n-r(X)]$ is the *mean square for subplot error* $[MSE(s)]$.

The results in Section 1 are for functions of γ that are estimable in model (11.1.2). We now show that $\lambda'\gamma$ is estimable in (11.1.2) if and only if $\lambda'\gamma$ is estimable in (2). The argument is given for real-valued estimable functions, but it clearly applies to vector-valued estimable functions. First, suppose that $\lambda'\gamma$ is estimable in (11.1.2). Then there exists a vector ξ such that $\lambda' = \xi'X_2$ and $\xi'X_1 = 0$. It follows immediately that $\lambda' = \xi'X_2$ and $\xi'X_* = 0$, so $\lambda'\gamma$ is estimable in model (2).

Now suppose that $\lambda'\gamma$ is estimable in model (2). There exists a vector ξ such that $\xi'X_* = 0$ and $\xi'X_2 = \lambda'$. Using equation (4) and $\xi'M_* = 0$, it is easily seen that

$$\lambda' = \xi'X_2 = \xi'M_ZX_2 = \xi'M_2X_2$$

and, since $M_2X_1 = 0$,

$$\xi' M_2 X_1 = 0.$$

The vector $\xi' M_2$ satisfies the two conditions needed to show that $\lambda'\gamma$ is estimable in (11.1.2). Thus, inferences about estimable functions $\lambda'\gamma$ can be made exactly as in Section 1.

Drawing inferences about δ is trickier. The projection operator M_Z is the sum of two orthogonal projection operators M_* and M_2. Estimation of $\lambda'\gamma$ can be accomplished easily because the estimate depends on $M_2 Y$ alone. Similarly, estimable functions whose estimates depend on $M_* Y$ alone can be handled simply. The problem lies in identifying which estimable functions have estimates that depend on $M_* Y$ alone. Since estimable functions of γ depend on $M_2 Y$, any estimable function with an estimate that depends on $M_* Y$ alone must involve δ. (Of course, there exist estimable functions that depend on both $M_* Y$ and $M_2 Y$.) Later, a condition will be discussed that forces all estimable functions of δ to depend only on $M_* Y$. With this condition, we have a convenient dichotomy in that estimates of functions of δ depend on the perpendicular projection operator M_*, and estimates of functions of γ depend on the perpendicular projection operator M_2. The condition referred to is convenient, but it is not necessary for having a generalized split plot model.

As discussed in Chapter 3 the question of whether the estimate of an estimable function, say $\Lambda'\beta_*$, depends only on $M_* Y$ is closely related to the constraint on the model imposed by the hypothesis $\Lambda'\beta_* = 0$. In particular, if $\Lambda' = P'Z$, then the constraint imposed by $\Lambda'\beta_* = 0$ is $\mathrm{E}(Y) \perp C(M_Z P)$, and $\Lambda'\hat{\beta}_*$ depends on $M_* Y$ if and only if $C(M_Z P) \subset C(M_*) = C(X_*)$. *In the discussion that follows,* $\Lambda'\beta_* = 0$ *is assumed to put a constraint on* $C(X_*)$.

We seek to derive an F test for $\Lambda'\beta_* = 0$. From Corollary 3.8.3,

$$\left(\Lambda'\hat{\beta}_*\right)' \left[\mathrm{Cov}\left(\Lambda'\hat{\beta}_*\right)\right]^- \left(\Lambda'\hat{\beta}_*\right) \sim \chi^2\left(r(\Lambda), \left(\Lambda'\beta_*\right)' \left[\mathrm{Cov}\left(\Lambda'\hat{\beta}_*\right)\right]^- \left(\Lambda'\beta_*\right)/2\right).$$

We need the covariance of $\Lambda'\hat{\beta}_*$. Note that $\Lambda'\hat{\beta}_* = P'M_Z Y = P'M_* Y$. From equation (1) and the fact that $C(X_*) \subset C(X_1)$, it is easily seen that

$$M_* V = [(1-\rho) + m\rho] M_*; \tag{6}$$

so

$$\begin{aligned} \mathrm{Cov}\left(\Lambda'\hat{\beta}_*\right) &= \sigma^2 P' M_* V M_* P = \sigma^2 [(1-\rho) + m\rho] P' M_* P \\ &= \sigma^2 [(1-\rho) + m\rho] P' M_Z P \\ &= \sigma^2 [(1-\rho) + m\rho] \Lambda' \left(Z'Z\right)^- \Lambda, \end{aligned}$$

which, except for the term $(1-\rho) + m\rho$, is the usual covariance of $\Lambda'\hat{\beta}_*$ from ordinary least squares. We can get an F test of $\Lambda'\beta_* = 0$ if we can find an independent chi-squared estimate of $\sigma^2 [(1-\rho) + m\rho]$.

Theorem 11.2.5. Under model (2),

$$Y'(M_1 - M_*)Y/\sigma^2[(1-\rho)+m\rho] \sim \chi^2(r(X_1) - r(X_*), 0).$$

PROOF. Observe that

$$M_1 V = [(1-\rho)+m\rho]M_1. \tag{7}$$

Using equations (6) and (7), it is easy to check the conditions of Theorem 1.3.6. It remains to show that $\beta_*' Z'(M_1 - M_*)Z\beta_* = 0$. Recall that $M_Z = M_* + M_2$, and note that $M = M_1 + M_2$. It follows that $M_1 - M_* = M - M_Z$. Clearly, $(M - M_Z)Z\beta_* = 0$. □

The quadratic form $Y'(M_1 - M_*)Y$ is called the *sum of squares for whole plot (cluster) error*. This is denoted $SSE(w)$. An unbiased estimate of $\sigma^2[(1-\rho)+m\rho]$ is available from

$$MSE(w) = Y'(M_1 - M_*)Y \big/ [r(X_1) - r(X_*)].$$

To complete the derivation for the F test of $\Lambda'\beta_* = 0$, we need to show that $\Lambda'\hat{\beta}_*$ and $Y'(M_1 - M_*)Y$ are independent. It suffices to note that

$$\begin{aligned} \operatorname{Cov}(M_*Y, (M_1 - M_*)Y) &= \sigma^2 M_* V(M_1 - M_*) \\ &= \sigma^2[(1-\rho)+m\rho]M_*(M_1 - M_*) \\ &= 0. \end{aligned}$$

The F test is based on the distributional result

$$\frac{\left(\Lambda'\hat{\beta}_*\right)'[\Lambda'(Z'Z)^-\Lambda]^-\left(\Lambda'\hat{\beta}_*\right)/r(\Lambda)}{MSE(w)} \sim F(r(\Lambda), r(X_1) - r(X_*), \pi),$$

where

$$\pi = \left(\Lambda'\beta_*\right)'\left[\Lambda'(Z'Z)^-\Lambda\right]^-\left(\Lambda'\beta_*\right)/2\sigma^2[(1-\rho)+m\rho]$$

and $(\Lambda'\beta_*)'[\Lambda'(Z'Z)^-\Lambda]^-(\Lambda'\beta_*) = 0$ if and only if $\Lambda'\beta_* = 0$.

The argument establishing the independence of $\Lambda'\hat{\beta}_*$ and $MSE(w)$ can be extended to establish the independence of all the distinct statistics being used.

Theorem 11.2.6. M_*Y, M_2Y, $SSE(w)$, and $SSE(s)$ are mutually independent.

PROOF. Since the joint distribution of Y is multivariate normal, it suffices to use equations (6) and (7) to establish that the covariance between any pair of M_*Y, M_2Y, $(M_1 - M_*)Y$, and $(I - M)Y$ is 0. □

To summarize the results so far, if the linear model satisfies Conditions 11.2.1 and 11.2.3, then (a) least squares estimates are BLUEs, (b) inferences about estimable functions $\lambda'\gamma$ can be made in the usual way (i.e., just as if $V = I$) with the exception that the estimate of error is taken to be $MSE(s)$, and (c) inferences about estimable functions of $\Lambda'\beta_*$ that put a constraint on $C(X_*)$ can be drawn in the usual way, except that $MSE(w)$ is used as the estimate of error.

As mentioned, it is not clear what kind of estimable functions put a constraint on $C(X_*)$. Two ways of getting around this problem will be discussed. As mentioned above, one way is to place another condition on the model matrix Z of model (2), a condition that forces the estimable functions of δ to put constraints on $C(X_*)$. A second approach, that requires no additional conditions, is to abandon the idea of testing estimable functions and to look at testing models.

Inferences About δ

One of the problems with generalized split plot models is in identifying the hypotheses that put constraints on $C(X_*)$. In general, such hypotheses can involve both the δ and the γ parameters. For example, suppose that $C(X_*) \cap C(X_2)$ contains a nonzero vector ξ. Then, since $M_Z\xi = \xi \in C(X_*)$, the hypothesis $\xi'X_*\delta + \xi'X_2\gamma = 0$ puts a constraint on $C(X_*)$. However, $\xi'X_*\delta + \xi'X_2\gamma$ involves both the δ and γ parameters because, with $\xi \in C(X_*) \cap C(X_2)$, neither $\xi'X_*$ nor $\xi'X_2$ is 0. The most common example of this phenomenon occurs when X_* and X_2 are chosen so that $C(X_*)$ and $C(X_2)$ both contain a column of 1s (i.e., J_n). It follows that inferences about the grand mean, $n^{-1}J_nZ\beta_*$, are made using $MSE(w)$.

The condition stated below ensures that any estimable function of the δs puts a constraint on $C(X_*)$. This condition is typically satisfied when Z is the model matrix for a balanced multifactor ANOVA (with some of the interactions possibly deleted).

Condition 11.2.7. For $v \in C(X)$, if $v \perp C(X_2)$, then $v \in C(X_1)$.

Suppose that $\lambda'\delta$ is an estimable function. Then $\lambda' = \xi'X_*$ and $\xi'X_2 = 0$ for some $\xi \in C(Z)$. Since $\xi'X_2 = 0$, Condition 11.2.7 implies that $\xi \in C(X_1)$; thus $M_2\xi = 0$. Finally, by (4),

$$M_Z\xi = M_*\xi \in C(X_*).$$

Thus, estimable functions of δ put a constraint on $C(X_*)$ and inferences about such functions are made using M_*Y and $MSE(w)$.

Testing Models

We will now examine the problem of testing model (2) against reduced models. To look at the complete problem, we will discuss both reduced models that put a constraint on $C(X_*)$ and reduced models that put a constraint on $C(X_2)$. For both kinds of reduced models, the tests are analogous to those developed in Section 3.2. A reduced model that puts a constraint on $C(X_*)$ can be tested by comparing the $SSE(w)$ for model (2) with the $SSE(w)$ for the reduced model. The difference in $SSE(w)$s is divided by the difference in the ranks of the design matrices to give a numerator mean square for the test. The denominator mean square is $MSE(w)$ from model (2). The test for a reduced model that puts a constraint on $C(X_2)$ is performed in a similar fashion using $SSE(s)$ and $MSE(s)$. In the discussion below, specific models and notation are presented to justify these claims.

First, consider a reduced model that puts a constraint on $C(X_*)$. The reduced model is a model of the form

$$Y = Z_0\xi + e, \quad Z_0 = [X_{0*}, X_2], \quad C(X_{0*}) \subset C(X_*). \tag{8}$$

Let $M_0 = Z_0(Z_0'Z_0)^- Z_0'$ and $M_{0*} = X_{0*}(X_{0*}'X_{0*})^- X_{0*}'$. If the equivalent of Condition 11.2.3 holds for model (8), then

$$\frac{Y'(M_* - M_{0*})Y/[r(X_*) - r(X_{0*})]}{MSE(w)} \sim F\left(r(X_*) - r(X_{0*}), r(X_1) - r(X_*), \pi\right),$$

where

$$\pi = \beta_*'Z'(M_* - M_{0*})Z\beta_*/2\sigma^2[(1-\rho) + m\rho]$$

and $\beta_*'Z'(M_* - M_{0*})Z\beta_* = 0$ if and only if $\mathrm{E}(Y) \in C(Z_0)$.

These results follow from Theorems 11.2.5 and 11.2.6 and Corollary 3.8.3 upon noticing two things: First, in Corollary 3.8.3, $A - A_0 = M_Z - M_0 = M_* - M_{0*}$. Second, from equation (6), $M_* = [(1-\rho) + m\rho]M_*V^{-1}$, with a similar result holding for $M_{0*}V^{-1}$.

The other kind of reduced model that can be treated conveniently is a reduced model that puts a constraint on $C(X_2)$. The reduced model is written as

$$Y = Z_0\xi + e, \quad Z_0 = [X_*, X_3], \quad C(X_3) \subset C(X_2). \tag{9}$$

If the equivalent of Condition 11.2.3 holds for model (9), write M_0 as before and $M_3 = (I - M_1)X_3\left[X_3'(I - M_1)X_3\right]^- X_3'(I - M_1)$. Then

$$\frac{Y'(M_2 - M_3)Y/[r(M_2) - r(M_3)]}{MSE(s)} \sim F\left(r(M_2) - r(M_3), n - r(X), \pi\right),$$

where

$$\pi = \beta_*'Z'(M_2 - M_3)Z\beta_*/2\sigma^2(1-\rho)$$

and $\beta_*'Z'(M_2 - M_3)Z\beta_* = 0$ if and only if $\mathrm{E}(Y) \in C(Z_0)$.

These results follow from Theorem 11.2.6, Corollary 3.8.3, and relation (5) upon noticing that $A - A_0 = M_Z - M_0 = M_2 - M_3$; and, since $M_2V = (1-\rho)M_2$, we have $(1-\rho)^{-1}M_2 = M_2V^{-1}$, and a similar result for M_3.

Identifying Generalized Split Plot Models

There are only two conditions necessary for having a generalized split plot model, Condition 11.2.1 and Condition 11.2.3. The form of generalized split plot models can be read from these conditions. Condition 11.2.1 requires that an equal number of observations be obtained within each whole plot. Condition 11.2.3 requires $C(Z) = C(X_*, (I-M_1)X_2)$, where $C(X_*) \subset C(X_1)$. Since $C(X_1)$ is the column space that allows a separate effect for each cluster, $X_*\delta$ can be anything that treats all of the observations in a given whole plot the same. The matrix X_2 can contain the columns for any ANOVA effects that are balanced within whole plots. X_2 can also contain any columns that are orthogonal to $C(X_1)$. Model (3) in Example 11.2.4 displays these characteristics.

Computations

The simplest way to actually fit generalized split plot models would seem to be to fit both models (2) and (11.1.2) using an ordinary least squares computer program. Fitting model (2) provides least squares estimates of δ and γ. Fitting model (2) also provides $Y'(I-M_Z)Y$ as the reported SSE. This reported SSE is not appropriate for any inferences, but, as seen below, it can be used to obtain the whole plot sum of squares error. Fitting model (11.1.2) provides least squares estimates of α and γ and the reported SSE is $Y'(I-M)Y$. If the model is a generalized split plot model, the two estimates of γ should be identical. Since $Y(I-M)Y$ is the $SSE(s)$ (sum of squares for subplot error), any conclusions about γ obtained from fitting (11.1.2) will be appropriate. To obtain $SSE(w)$ (sum of squares for whole plot error), note that

$$\begin{aligned} Y'(I-M_Z)Y - Y'(I-M)Y &= Y'(M-M_Z)Y \\ &= Y'(M_1+M_2-M_*-M_2)Y \\ &= Y'(M_1-M_*)Y. \end{aligned}$$

Thus, all of the computationally intensive work can be performed on standard computer programs.

Exercise 11.8 Give detailed proofs of the test statistic's distribution for

(a) testing model (8) against model (2),

(b) testing model (9) against model (2).

11.3 The Split Plot Design

The traditional model for a split plot design is a special case of the model presented in Section 2. We will present the split plot model and a model equivalent to model (11.1.2). We will use the balance of the split plot design to argue that Conditions 11.2.1, 11.2.3, and 11.2.7 hold. The arguments based on the balance of the split plot model are similar to those presented in Section 7.6. Statistical inferences are based on least squares estimates and quadratic forms in corresponding perpendicular projection matrices. The balance of the split plot model dictates results that are very similar to those described in Sections 7.1, 7.2, and 7.6.

The traditional split plot model involves a randomized complete block design in the whole plots. In fact, a completely randomized design or a Latin square in the whole plots leads to an analogous analysis. Suppose that there are r blocks of t whole plots available. Within each block, a different (whole plot) treatment is applied to each whole plot. Let μ denote a grand mean, ξs denote block effects, and ωs denote whole plot treatment effects. Let each whole plot be divided into m subplots with a different (subplot) treatment applied to each subplot. The τs denote subplot treatment effects, and the $(\omega\tau)$s denote interaction effects between the whole plot treatments and the subplot treatments. The split plot model has two sources of error, whole plot to whole plot variation denoted by η, and subplot to subplot variation denoted by e. The split plot model is

$$y_{ijk} = \mu + \xi_i + \omega_j + \eta_{ij} + \tau_k + (\omega\tau)_{jk} + e_{ijk}, \tag{1}$$

$i = 1,\ldots,r$, $j = 1,\ldots,t$, $k = 1,\ldots,m$, η_{ij}s independent $N(0,\sigma_w^2)$, e_{ijk}s independent $N(0,\sigma_s^2)$. The η_{ij}s and e_{ijk}s are assumed to be independent. We can combine the error terms as $\varepsilon_{ijk} = \eta_{ij} + e_{ijk}$. Writing the vector of errors as $\varepsilon = [\varepsilon_{111},\varepsilon_{112},\ldots,\varepsilon_{rtm}]'$, we get

$$\begin{aligned}\mathrm{Cov}(\varepsilon) &= \mathrm{Blk\ diag}[\sigma_s^2 I_m + \sigma_w^2 J_m^m] \\ &= \sigma^2[(1-\rho)I + m\rho M_1],\end{aligned}$$

where $\sigma^2 = \sigma_w^2 + \sigma_s^2$ and $\rho = \sigma_w^2/(\sigma_w^2 + \sigma_s^2)$. In a split plot model, the whole plots are considered as the different combinations of i and j. There are m observations for each whole plot, so Condition 11.2.1 holds.

The model that includes separate effects α_{ij} for each whole plot can be written as

$$y_{ijk} = \alpha_{ij} + \eta_{ij} + \tau_k + (\omega\tau)_{jk} + e_{ijk}\,.$$

Combining the error terms gives

$$y_{ijk} = \alpha_{ij} + \tau_k + (\omega\tau)_{jk} + \varepsilon_{ijk}\,,$$

or, using a parameterization with interactions,

$$y_{ijk} = \mu + \xi_i + \omega_j + (\xi\omega)_{ij} + \tau_k + (\omega\tau)_{jk} + \varepsilon_{ijk}\,. \tag{2}$$

From Section 2, $C(X_1)$ is the space spanned by the columns associated with the α_{ij}s and $C(X_2)$ is the space spanned by the columns associated with the τ_ks and $(\omega\tau)_{jk}$s. $C(X)$ is the column space for model (2). Using the parameterization of model (2) and the notation and results of Section 7.6 gives

$$\begin{aligned} C(X_1) &= C(M_\mu + M_\xi + M_\omega + M_{\xi\omega}), \\ C(X_2) &= C(M_\mu + M_\omega + M_\tau + M_{\omega\tau}), \\ C(X) &= C(M_\mu + M_\xi + M_\omega + M_{\xi\omega} + M_\tau + M_{\omega\tau}). \end{aligned}$$

Recall that all of the M matrices on the right-hand sides are perpendicular projection matrices, and that all are mutually orthogonal. In particular, $M = M_\mu + M_\xi + M_\omega + M_{\xi\omega} + M_\tau + M_{\omega\tau}$ and $M_1 = M_\mu + M_\xi + M_\omega + M_{\xi\omega}$.

The split plot model (1) is a reduced model relative to model (2). The $\xi\omega$ interactions are dropped to create the split plot model. $C(X_*)$ is the space spanned by the columns associated with μ, the ξ_is, and the ω_js. Again using results from 7.6,

$$\begin{aligned} C(X_*) &= C(M_\mu + M_\xi + M_\omega), \\ C[(I - M_1)X_2] &= C(M_\tau + M_{\omega\tau}), \\ C(Z) &= C(M_\mu + M_\xi + M_\omega + M_\tau + M_{\omega\tau}). \end{aligned}$$

Clearly, Condition 11.2.3 applies.

In fact, even Condition 11.2.7 holds, so that estimable functions of the ξs and ωs are tested using $MSE(w)$. To check Condition 11.2.7, it suffices to show that if $v \in C(X)$ and $v \perp C(X_2)$, then $v \in C(X_1)$. If $v \in C(X)$, then $Mv = v$. If $v \perp C(X_2)$, then $(M_\mu + M_\omega + M_\tau + M_{\omega\tau})v = 0$. Thus

$$\begin{aligned} v = Mv &= (M_\mu + M_\xi + M_\omega + M_{\xi\omega} + M_\tau + M_{\omega\tau})v \\ &= (M_\xi + M_{\xi\omega})v. \end{aligned}$$

But, $v = (M_\xi + M_{\xi\omega})v \in C(X_1)$. It should be noted that with $(\omega\tau)$ interaction in the model, contrasts in the ωs are not estimable. The usual procedure gives estimates of contrasts in the $\omega_j + \overline{(\omega\tau)}_{j\cdot}$s. Without $(\omega\tau)$ interaction, contrasts in the ωs become estimable. In either case, the estimates are obtained using M_*.

We can now write out an ANOVA table.

Source	df	SS	E(MS)
μ	$r(M_\mu)$	$Y'M_\mu Y$	$(\sigma_s^2 + m\sigma_w^2) + \beta' X' M_\mu X\beta / r(M_\mu)$
ξ	$r(M_\xi)$	$Y'M_\xi Y$	$(\sigma_s^2 + m\sigma_w^2) + \beta' X' M_\xi X\beta / r(M_\xi)$
ω	$r(M_\omega)$	$Y'M_\omega Y$	$(\sigma_s^2 + m\sigma_w^2) + \beta' X' M_\omega X\beta / r(M_\omega)$
error 1	$r(M_{\xi\omega})$	$Y'M_{\xi\omega}Y$	$\sigma_s^2 + m\sigma_w^2$
τ	$r(M_\tau)$	$Y'M_\tau Y$	$\sigma_s^2 + \beta' X' M_\tau X\beta / r(M_\tau)$
$\omega\tau$	$r(M_{\omega\tau})$	$Y'M_{\omega\tau}Y$	$\sigma_s^2 + \beta' X' M_{\omega\tau} X\beta / r(M_{\omega\tau})$
error 2	$r(I - M)$	$Y'(I - M)Y$	σ_s^2
Total	n	$Y'Y$	

Note that $\sigma^2[(1-\rho)+m\rho] = \sigma_s^2 + m\sigma_w^2$ and $\sigma^2(1-\rho) = \sigma_s^2$. Algebraically, we can write the table as

Source	df	SS	E(MS)
μ	1	$rtm\bar{y}_{...}^2$	
ξ	$r-1$	$tm\sum_i(\bar{y}_{i..}-\bar{y}_{...})^2$	$(\sigma_s^2+m\sigma_w^2)+A$
ω	$t-1$	$rm\sum_j(\bar{y}_{.j.}-\bar{y}_{...})^2$	$(\sigma_s^2+m\sigma_w^2)+B$
error 1	$(r-1)(t-1)$	$m\sum_{ij}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$	$\sigma_s^2+m\sigma_w^2$
τ	$m-1$	$rt\sum_k(\bar{y}_{..k}-\bar{y}_{...})^2$	σ_s^2+C
$\omega\tau$	$(t-1)(m-1)$	$r\sum_{jk}(\bar{y}_{.jk}-\bar{y}_{.j.}-\bar{y}_{..k}+\bar{y}_{...})^2$	σ_s^2+D
error 2	by subtraction	by subtraction	σ_s^2
Total	n	$\sum_{ijk}y_{ijk}^2$	

$$A = tm\sum_i(\xi_i-\bar{\xi}_\cdot)^2/(r-1),$$

$$B = rm\sum_j(\omega_j+\overline{(\omega\tau)}_{j\cdot}-\bar{\omega}_\cdot-\overline{(\omega\tau)}_{\cdot\cdot})^2/(t-1),$$

$$C = rt\sum_k(\tau_k+\overline{(\omega\tau)}_{\cdot k}-\bar{\tau}_\cdot-\overline{(\omega\tau)}_{\cdot\cdot})^2/(m-1),$$

$$D = r\sum_{jk}((\omega\tau)_{jk}-\overline{(\omega\tau)}_{j\cdot}-\overline{(\omega\tau)}_{\cdot k}+\overline{(\omega\tau)}_{\cdot\cdot})^2/(t-1)(m-1).$$

Tests and confidence intervals for contrasts in the τ_ks and $(\omega\tau)_{jk}$s are based on the usual least squares estimates and use the mean square from the "error 2" line for an estimate of variance. Tests and confidence intervals in the ξ_is and ω_js also use least squares estimates, but use the mean square from the "error 1" line for an estimate of variance. Note that contrasts in the ω_js are really contrasts in the $(\omega_j+\overline{(\omega\tau)}_{j\cdot})$s when interaction is present.

Finally, a word about missing data. If one or more whole plots are missing, the data can still be analyzed as in Section 2. If one or more subplots are missing, the data can still be analyzed as in Section 1; however, with missing subplots, some sort of ad hoc analysis for the whole plot effects must be used.

Exercise 11.9 Consider the table of means

		τ			
		1	2	$\cdots$	m
ω	1	$\bar{y}_{\cdot 11}$	$\bar{y}_{\cdot 12}$	$\cdots$	$\bar{y}_{\cdot 1m}$
	2	$\bar{y}_{\cdot 21}$	$\bar{y}_{\cdot 22}$	$\cdots$	$\bar{y}_{\cdot 2m}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	t	$\bar{y}_{\cdot t1}$	$\bar{y}_{\cdot t2}$	$\cdots$	$\bar{y}_{\cdot tm}$

Let $\sum_{k=1}^m d_k = 0$. For any fixed j, find a confidence interval for $\sum_{k=1}^m d_k\mu_{jk}$, where $\mu_{jk} = \mu+\bar{\xi}_\cdot+\omega_j+\tau_k+(\omega\tau)_{jk}$. (Hint: The estimate of the variance comes from the

"error 2" line.) Let $\sum_{j=1}^{t} c_j = 0$. Why is it not possible to find a confidence interval for $\sum_{j=1}^{t} c_j \mu_{jk}$?

11.4 Identifying the Appropriate Error

Statistics is all about drawing conclusions from data that are subject to error. One of the crucial problems in statistics is identifying and estimating the appropriate error so that valid conclusions can be made. The importance of this issue has long been recognized by the statistical community. The necessity of having a valid estimate of error is one of the main points in *The Design of Experiments*, Fisher's (1935) seminal work.

The key feature of split plot models is that they involve two separate sources of variation. The analysis involves two separate estimates of error, and a correct analysis requires that the estimates be used appropriately. If the existence of two separate sources of variability is not noticed, the estimate of error will probably be obtained by pooling the sums of squares for the two separate errors. The pooled estimate of error will generally be too small for comparing treatments applied to whole plots and too large for comparing treatments applied to subplots. The whole plot treatments will appear more significant than they really are. The subplot treatments will appear less significant than they really are. The interactions between whole plot and subplot treatments will also appear less significant than they really are.

The problem of identifying the appropriate error is a difficult one. In this section, some additional examples of the problem are discussed. First, the problem of subsampling is considered. The section ends with a discussion of the appropriate error for testing main effects in the presence of interactions.

Subsampling

One of the most commonly used, misused, and abused of models is the subsampling model.

EXAMPLE 11.4.1. In an agricultural experiment, one treatment is applied to each of 6 pastures. The experiment involves 4 different treatments, so there are a total of 24 pastures in the experiment. On each pasture 10 observations are taken. The 10 observations taken on each pasture are referred to as subsamples. Each observation is subject to two kinds of variability: (1) pasture to pasture variability and (2) within pasture variability. Note, however, that in comparing the 10 observations taken on a given pasture, there is no pasture to pasture variability. The correct model for this experiment involves error terms for both kinds of variability. Typically the model is taken as

$$y_{ijk} = \mu + \omega_i + \eta_{ij} + e_{ijk}, \tag{1}$$

$i = 1,2,3,4$, $j = 1,\ldots,6$, $k = 1,\ldots,10$, η_{ij}s i.i.d. $N(0,\sigma_w^2)$, e_{ijk}s i.i.d. $N(0,\sigma_s^2)$, and the η_{ij}s and e_{ijk}s are independent. In this model, σ_w^2 is the pasture to pasture variance and σ_s^2 is the within pasture or subsampling variance.

As will be seen below, the statistical analysis of model (1) acts like there is only 1 observation on each pasture. That 1 observation is the mean of the 10 actual observations that were taken. If the analysis acts like only 1 observation was taken on a pasture, why should an experimenter trouble to take 10 observations? Why not take just 1 observation on each pasture?

With 1 real observation on each pasture, the statistical analysis is subject to the whole weight of the within pasture variability. By taking 10 observations on a pasture and averaging them to perform the analysis, the mean of the 10 observations still has the full pasture to pasture variability, but the within pasture variability (variance) is cut by a factor of 10. The experiment is being improved by reducing the variability of the treatment estimates, but that improvement comes only in the reduction of the within pasture variability. The effects of pasture to pasture variability are not reduced by subsampling.

Rather than subsampling, it would be preferable to use more pastures. Using more pastures reduces the effects of both kinds of variability. Unfortunately, doing that is often not feasible. In the current example, the same reduction in within pasture variation without subsampling would require the use of 240 pastures instead of the 24 pastures that are used in the experiment with subsampling. In practice, obtaining 24 pastures for an experiment can be difficult. Obtaining 240 pastures can be well nigh impossible.

In general, a subsampling model is

$$y_{ijk} = \mu + \omega_i + \eta_{ij} + e_{ijk}, \quad \mathrm{Var}(\eta_{ij}) = \sigma_w^2, \quad \mathrm{Var}(e_{ijk}) = \sigma_s^2, \tag{2}$$

$i = 1,\ldots,t$, $j = 1,\ldots,r$, $k = 1,\ldots,m$. By checking Conditions 11.2.1, 11.2.3, and 11.2.7, it is easily seen that this is a generalized split plot model with X_2 vacuous. An ANOVA table can be written:

Source	df	SS	E(MS)
ω	$t-1$	$rm\sum_i(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2$	$\sigma_s^2 + m\sigma_w^2 + A$
error 1	$t(r-1)$	$m\sum_{ij}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2$	$\sigma_s^2 + m\sigma_w^2$
error 2	$rt(m-1)$	$\sum_{ijk}(y_{ijk} - \bar{y}_{ij\cdot})^2$	σ_s^2

where $A = rm\sum_i(\omega_i - \bar{\omega}_\cdot)^2/(t-1)$. The entire analysis is performed as a standard one-way ANOVA. The variance of $\bar{y}_{i\cdot\cdot}$ is $(\sigma_s^2 + m\sigma_w^2)/rm$, so the "error 1" line is used for all tests and confidence intervals.

As mentioned above, an equivalent analysis can be made with the averages of the observations in each subsample. The model based on the averages is

$$\bar{y}_{ij\cdot} = \mu + \omega_i + \xi_{ij}, \quad \mathrm{Var}(\xi_{ij}) = [\sigma_w^2 + (\sigma_s^2/m)],$$

$i = 1,\ldots,t,\ j = 1,\ldots,r$. It is easily seen that this gives exactly the same tests and confidence intervals as those obtained from model (2).

One of the most common mistakes in statistical practice is to mistake subsampling for independent replication in an experiment. Example 11.4.1 involves 6 independent replications, i.e., the 6 pastures to which each treatment is applied. The 10 observations on a given pasture are not independent because the random effect for pastures is the same for all of them. The incorrect model that is often analyzed instead of model (2) is

$$y_{ij} = \mu + \omega_i + e_{ij}, \tag{3}$$

$i = 1,\ldots,t,\ j = 1,\ldots,rm$. The effect of analyzing model (3) is that the "error 1" and "error 2" lines of the ANOVA are pooled together. Since the expected mean square for "error 2" is only σ_s^2, the pooled mean square error is inappropriately small and all effects will appear to be more significant than they really are.

Subsampling can be an important tool, especially when the variability between subsamples is large. However, it is important to remember that subsampling is to be used in addition to independent replication. It does not eliminate the need for an adequate number of independent replicates. In terms of Example 11.4.1, an experimenter should first decide on a reasonable number of pastures and then address the question of how many observations to take within a pasture. If the pasture to pasture variability is large compared to the within pasture variability, subsampling will be of very limited value. If the within pasture variability is large, subsampling can be extremely worthwhile.

The existence of subsampling can usually be determined by carefully identifying the treatments and the experimental units to which the treatments are applied. In the agricultural example, identifying the subsampling structure was easy. Treatments were applied to pastures. If differences between treatments are to be examined, then differences between pastures with the same treatment must be error.

Lest the reader think that identifying subsampling is easy, let us try to confuse the issue. The analysis that has been discussed is based on the assumption that pasture to pasture variability is error, but now suppose that the experimenter has an interest in the pastures. For example, different pastures have different fertilities, so some pastures are better than others. If differences in pastures are of interest, it may be reasonable to think of the η_{ij}s as fixed effects, in which case there is no subsampling and the ANOVA table gives only one error line. The expected mean squares are:

Source	df	E(MS)
ω	3	$\sigma^2 + 20\sum_i(\omega_i + \bar{\eta}_{i\cdot} - \bar{\omega}_{\cdot} - \bar{\eta}_{\cdot\cdot})^2$
η	20	$\sigma^2 + .5\sum_{ij}(\eta_{ij} - \bar{\eta}_{i\cdot})^2$
error	216	σ^2

The mean square for η can be used to test whether there are any differences between pastures that have the same treatment. The $MSE(\omega)$ provides a test of whether the treatment effects added to their average pasture effects are different. The test from

$MSE(\omega)$ may not be very interesting if there are different pasture effects. In summary, the analysis depends crucially on whether the η_{ij}s are assumed to be random or fixed. When the ω treatments are of primary importance it makes sense to treat the η effects as random.

Two-Way ANOVA with Interaction

The two-way ANOVA with interaction model from Section 7.2 is

$$y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + e_{ijk}, \tag{4}$$

$i = 1,\ldots,a$, $j = 1,\ldots,b$, $k = 1,\ldots,N$. First, the question of subsampling in a two-way ANOVA will be addressed. The discussion of subsampling leads to an examination of independent replication, and to the question of whether the interactions should be considered fixed or random. The discussion of identifying the appropriate error begins with an example:

EXAMPLE 11.4.2. We want to investigate the effects of 4 fertilizers and 6 herbicides on pastures used for raising cane (as in sugar cane). There are a total of $4 \times 6 = 24$ treatment combinations. If each treatment combination is applied to 5 pastures, then model (4) is appropriate with $a = 4$, $b = 6$, and $N = 5$.

Now consider an alternative experimental design that is easily confused with this. Suppose each treatment combination is applied to 1 pasture and 5 observations are taken on each pasture. There is no independent replication. The 5 observations on each pasture are subsamples. Comparisons within pastures do not include pasture to pasture variability. If model (4) is used to analyze such data, the *MSE* is actually the estimated subsampling variance. It is based on comparisons within pastures. An analysis based on model (4) will inflate the significance of all effects.

The appropriate model for this subsampling design is

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

where $\xi_{ij} \sim N(0, \sigma_w^2)$ and $e_{ijk} \sim N(0, \sigma_s^2)$. Note that the indices on γ and ξ are exactly the same. It is impossible to tell interactions apart from pasture to pasture variability. As a result, unless the interactions can be assumed nonexistent, there is no appropriate estimate of error available in this experiment.

In the example, two extreme cases were considered, one case with no subsampling and one case with no independent replication. Of course, any combination of subsampling and independent replication can be used. In the example, the designs were clearly stated so that the subsampling structures were clear. In practice, this rarely occurs. When presented with data that have been collected, it can be very difficult to identify how the experiment was designed.

Designs without independent replications are actually quite common. When confronted with a two-factor ANOVA without any independent replication, the fixed interaction effects are generally assumed to be zero so that an analysis can be performed. This is precisely the assumption made in Chapter 8 in analyzing the Randomized Complete Block model. An alternate way of phrasing this idea is that any interaction effects that exist must be due to error. This idea that interaction effects can themselves be errors is important. If the interactions are errors, then model (4) needs to be changed. The standard assumption would be that the γ_{ij}s are independent $N(0, \sigma_w^2)$ random variables.

Note that the assumption of random γ_{ij}s does not imply the existence of subsampling. Subsampling is a property of the experimental design. What is being discussed is simply a choice about how to model interactions. Should they be modeled as fixed effects, or should they be modeled as random errors? One guideline is based on the repeatability of the results. If the pattern of interactions should be the same in other similar studies, then the interactions are fixed. If there is little reason to believe that the pattern of interactions would be the same in other studies, then the interactions would seem to be random.

The analysis of model (4) with random interactions is straightforward. The model is a generalized split plot model with X_2 vacuous. The mean square for interactions becomes the mean square for "error 1." The mean square error from assuming fixed interactions becomes the mean square for "error 2." The analysis for main effects uses the mean square "error 1" exclusively as the estimate of error.

Although this is not a subsampling model, it does involve two sources of variation: 1) variation due to interactions, and 2) variation due to independent replication (i.e., variation from experimental unit to experimental unit). It seems to be difficult to reduce the effects on comparisons among treatments of the variability due to interactions. The effect of variation due to experimental units can be reduced by taking additional independent replications.

11.5 Exercise: An Unusual Split Plot Analysis

Cornell (1988) considered data on the production of vinyl for automobile seat covers. Different blends involve various plasticizers, stabilizers, lubricants, pigments, fillers, drying agents, and resins. The current data involve 5 blends of vinyl.

The 5 blends represent various percentages of 3 plasticizers that together make up 40.7% of the product. The first plasticizer is restricted to be between 19.1% and 34.6% of the product. The second plasticizer is restricted to be between 0% and 10.2% of the product. The third plasticizer is restricted to be between 6.1% and 11.4% of the product. Changing these restrictions to fractions of the 40.7% of the total, we get

$$0.47 \leq x_1 \leq 0.85, \quad 0 \leq x_2 \leq 0.25, \quad 0.15 \leq x_1 \leq 0.28.$$

The 5 blends are

Blend	(x_1, x_2, x_3)
1	(0.85, 0.000, 0.150)
2	(0.72, 0.000, 0.280)
3	(0.60, 0.250, 0.150)
4	(0.47, 0.250, 0.280)
5	(0.66, 0.125, 0.215)

Note that the first four blends have all combinations of x_2 and x_3 at their extremes with x_1 values decreasing at about 0.13 per blend. Blend 5 is in the center of the other blends. In particular, for $i = 1, 2, 3$, the x_i value of blend 5 is the mean of the other four x_i values. Eight groups of the five different blends were prepared.

The first group of 5 blends was run with the production process set for a high rate of extrusion ($z_1 = 1$) and a low drying temperature ($z_2 = -1$). The process was then reset for low extrusion rate and high drying temperature ($z_1 = -1, z_2 = 1$), and another group of 5 blends was run. For subsequent runs of 5 blends, the process was set for $z_1 = -1, z_2 = -1$, and $z_1 = 1, z_2 = 1$ to finish the first replication. Later, the second replication was run in the order $z_1 = -1, z_2 = 1$; $z_1 = 1, z_2 = 1$; $z_1 = 1, z_2 = -1$; $z_1 = -1, z_2 = -1$. The data are presented in Table 11.1

Table 11.1 Cornell's Scaled Vinyl Thickness Values.

Blend	x_1	x_2	x_3	z_1	z_2	Rep. 1	Rep. 2
1	0.85	0.000	0.150	1	−1	8	7
2	0.72	0.000	0.280	1	−1	6	5
3	0.60	0.250	0.150	1	−1	10	11
4	0.47	0.250	0.280	1	−1	4	5
5	0.66	0.125	0.215	1	−1	11	10
1	0.85	0.000	0.150	−1	1	12	10
2	0.72	0.000	0.280	−1	1	9	8
3	0.60	0.250	0.150	−1	1	13	12
4	0.47	0.250	0.280	−1	1	6	3
5	0.66	0.125	0.215	−1	1	15	11
1	0.85	0.000	0.150	−1	−1	7	8
2	0.72	0.000	0.280	−1	−1	7	6
3	0.60	0.250	0.150	−1	−1	9	10
4	0.47	0.250	0.280	−1	−1	5	4
5	0.66	0.125	0.215	−1	−1	9	7
1	0.85	0.000	0.150	1	1	12	11
2	0.72	0.000	0.280	1	1	10	9
3	0.60	0.250	0.150	1	1	14	12
4	0.47	0.250	0.280	1	1	6	5
5	0.66	0.125	0.215	1	1	13	9

Compare the RCB model for the five blend treatments

$$y_{hijk} = \gamma_h + \tau_{ijk} + \varepsilon_{hijk},$$

where the triples (i, j, k) only take on five distinct values, with the reduced regression model

$$y_{hijk} = \gamma_h + \beta_1 x_{1i} + \beta_2 x_{2j} + \beta_3 x_{3k} + \varepsilon_{hijk}.$$

Test lack of fit. Average over blends to do a whole plot analysis. Finally, do a split plot analysis.