NASA TECHNICAL
MEMORANDUM

NASA TM X-58030
JUNE 1969

A NEW APPROACH TO THE ANALYSIS OF LINEAR MODELS

By Fred Michael Speed

A Dissertation Submitted to
the Graduate College of Texas A & M University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Statistics

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

ABSTRACT

A New Approach to the Analysis of Linear Models
(May 1969)

Fred M. Speed, B.S., St. Mary's University
M.S., St. Mary's University

Directed by: Dr. R. R. Hocking

This dissertation analyzes linear models by writing the model in the form $Y = Xu + e$ subject to $\theta^T u = \xi$, where the usual assumptions are made about e. It is shown that for experimental design models, $X^T X$ is always diagonal and of full rank. Included are methods to obtain θ^T for the classical design models as well as arbitrary linear models. Other topics are regression models, covariance models, estimates of fixed effects in mixed models, and a procedure for obtaining expectations, variances and covariances for quadratic forms.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (concluded)

CHAPTER 1. INTRODUCTION

1.0 Preliminaries

The theory of linear models is one of the basic tools used by the statistician in analyzing data. It has evolved from a strictly algebraic tool, which could be used on hand calculators, to a sophisticated matrix technique which is amenable to high-speed computers.

The classical theory of linear models has as its foundation the works of R. A. Fisher [13]. Fisher's approach is based mainly on intuitive and heuristic concepts. He informs his readers that his methods are the logical way to analyze the data. Fisher states that his book is to be a handbook for research workers, expecially biologists. He makes no attempt to justify his results on a rigorous mathematical basis. The textbooks [14], [32], [34] published between 1915-1945, for the most part, follow this trend. They present the statistics in a "cookbook" manner; that is, if the experimenter follows the algebraic recipes given, he is

---

NOTE – The citations on the following pages follow the Journal of the American Statistical Association.

assured of obtaining the desired statistic. In 1947, Bartlet [2], Cochran [6], and Eisenhart [11] stressed the need of putting the basic concepts supporting the analysis of variance on a more mathematical foundation. From then until 1962, authors like Brownlee [4], Cramer [7], Scheffe [28], and Wilks [36] presented the mathematical approach to the analysis of variance. One disadvantage of their works, however, is that the analysis is derived using a strictly algebraic approach, and while their results are mathematically valid, they have lost their intuitive appeal.

About 1955, the elements of matrix theory were applied to the analysis of variance. Graybill [15], Rao [27], and Searle [31] published textbooks which use matrix theory as a basic tool. The result of using matrices is to provide not only a rigorous mathematical backing for the analysis of variance but also a conceptually concise theory of linear models, hence, a concise theory for the analysis of variance. Recently, several writers [20], [22] have been advocating using linear algebra in studying linear models. An appreciation of this method requires a detailed knowledge of linear algebra, function analysis, tensor products and projections.

This dissertation provides another improvement in that it provides a method of analysis that makes use of the very simple structure of the observations. That is, we assume that each observation came from a population with a mean and a variance. If we have more than one population, we may know certain relations about the means of these populations. All hypothesis testing and estimation is done in terms of the population means. It is shown in Chapter 3 that the classical linear models can be considered in this manner. Using this approach with matrix algebra, we have a theory of linear models that is both mathematically concise and conceptually clear from the viewpoint of statistics.

This approach avoids the misunderstanding caused by imposing "nonestimable" conditions on the nonfull rank models. It provides an understanding of using additional information in mixed models to obtain better estimates of the fixed effects. It avoids using special formulae when there are missing cells. There is no need to have one theory for the balanced case and another theory for the unbalanced case. This approach also suggests a technique, which is computationally efficient, to find expectations, variances and covariances of $Y^T AY$ and $Y^T BY$ where $Y \sim N(u,V)$, and V is not necessarily diagonal.

The value of this approach as a teaching tool is fourfold. First, the student uses fundamental concepts and relationships to derive the results. Second, the student is taught basic principles that apply both to the equal and unequal number per cell case. He is not taught a lot of "tricks" that apply only to the balanced case — "tricks" that would throw him into a quandary when he encounters missing cells, etc. Third, there is no question about what is being tested. The only way to obtain a sum of squares for an F test is to specify that the hypothesis to be tested is H_o: $\lambda^T u = \xi$, where λ^T is the hypothesis matrix, and ξ is known (usually zero). Finally, the interpretation of interactions and main effects has more meaning since these are related to means of populations and their interpretation in this light is clear.

As a computational tool, this approach is designed for high-speed computers. Since we have established the relation between this approach and the classical models, whenever we have a balanced experimental design model, we will use the special computational methods to obtain estimates or sum of squares and then continue the analysis from our viewpoint. A computer program, which was written in FORTRAN V, efficiently performs the analysis described in this dissertation.

The dissertation has seven chapters. This chapter contains the definitions and theorems necessary for understanding this approach. Chapter 2 provides the theory for analysis of the general "u" model. Chapter 3 provides the relationships between the classical models and the "u" model. Chapter 4 covers the analysis of mixed models. Chapter 5 provides analysis for regression models and covariance models. A discussion of tests for main effects and interactions also is included in Chapter 5. Chapter 6 gives several numerical examples illustrating the theory developed. And Chapter 7 considers some of the research that still is needed in this area.

1.1 Definitions and Theorems

In this section, we will give the definitions and theorems necessary to understand this approach. The definitions are all referenced. The first group of theorems, 1.1 to 1.15, is also found in the literature and has been referenced. The next set of theorems, 1.A to 1.F, has been proved by the author and the proofs are included.

<u>Definition 1</u> [15] – An $n \times n$ matrix A is said to be idempotent if $A \cdot A = A$.

<u>Definition 2</u> [15] – An $n \times n$ matrix P is said to be orthogonal if $P^T P = I$.

<u>Definition 3</u> [15] – If A is an $n \times n$ matrix and X and α are such that $AX = \alpha X$, then X is said to be an eigenvector of A and α the corresponding eigenvalue.

<u>Definition 4</u> [15] – When we write $Y^T AY$, we mean that A is symmetric.

We will assume that

$$Y = Xu + e$$

where, unless otherwise indicated,

X is a full rank matrix,

u is a constant vector, and

$e \sim N(0,\sigma^2 I)$.

We will write

$$\min_{\theta^T u=\delta} Q(u)$$

and mean that we want to minimize $Q(u)$ subject to $\theta^T u = \delta$. Also

$$\min_{\substack{\theta^T u=\delta \\ \Lambda^T u=\xi}} Q(u)$$

means that we want to minimize $Q(u)$ subject to $\theta^T u = \delta$ and $\Lambda^T u = \xi$.

<u>Theorem 1.1</u> [27] — Let A be an arbitrary $r\times p$ matrix. The four equations

(i) $AXA = A$

(ii) $XAX = X$

(iii) $(AX)^T = AX$

(iv) $(XA)^T = XA$

have a unique solution X . We call X the generalized inverse of A and denote X by A^+ .

If the vector equation $Ay = b$ is consistent, then the general solution is $y = A^+b + [I - A^+A]z$ where z is arbitrary. Note that $AA^+b = b$ and $A[I - A^+A] = 0$.

The vector equation $Ay = b$ is consistent if and only if $AA^{+}b = b$.

If A is $r\times p$ where $r < p$ and A is of rank r , then the rank of $I-A^{+}A$ is $p-r$.

Also note that $A^{+} = (A^{T}A)^{+}A^{T}$

<u>Theorem 1.2</u> [27] – Let $Y \sim N(u,\Sigma)$. If $Z = AY + \delta_0$, where δ_0 is fixed, then $Z \sim N(Au + \delta_0, A\Sigma A^{T})$.

<u>Theorem 1.3</u> [15] – Let $Y \sim N(u,\sigma^2 I)$, and let

$$\frac{Y^TY}{\sigma^2} = \sum_{i=1}^{k} \frac{Y^TA_iY}{\sigma^2} ,$$

where n_i is the rank of A_i . Any one of the three conditions listed below is a necessary and sufficient condition that the following two statements be true

(i) $Y^TA_iY/\sigma^2 \sim \chi^2(n_i,\lambda_i)$, where $\lambda_i = u^TA_iu$

(ii) Y^TA_iY, Y^TA_jY are independent if $i \neq j$.

The conditions are:

(1) A_i is idempotent for all $i = 1,\cdots,k$

(2) $A_iA_j = 0$ for all $i \neq j$

(3) $n = \Sigma n_i$

Theorem 1.4 [15] – If $Y \sim N(u,V)$, then $Y^T BY$ is distributed $\chi^2(k,\lambda)$, where $k = \text{rank } B$ and $\lambda = u^T Bu$, if and only if BV is idempotent. In the case where $V = \sigma^2 I$ and $Y^T BY = \frac{Y^T AY}{\sigma^2}$, the necessary and sufficient condition is that $A^2 = A$.

Theorem 1.5 [27] – Let A be an m×m symmetric matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, and $P_1 \cdots, P_m$ corresponding eigenvectors. Then

$$\max_x \frac{x^T Ax}{x^T x} = \lambda_1$$

or

$$\min_x \frac{x^T Ax}{x^T x} = \lambda_m \quad .$$

Theorem 1.6 [15] – If $T \sim \chi^2(p,\lambda)$, $Z \sim \chi^2(r)$ and T and Z are independent, then $u = \frac{rT}{pZ}$ is distributed as $F[p,r,\lambda]$.

Theorem 1.7 [27] – Let $R^2 = \min\, [Y - Xu]^T[Y - Xu]$ subject to $\Lambda^T u = \xi$ and where X is not necessarily of full rank, then

$$R^2 = R_0^2 + SS(\Lambda^T,\xi)$$

where

$$R_0^2 = Y^T Y - u^{*T}(X^T X)u^*$$

$$SS(\Lambda^T,\xi) = [\Lambda^T u^* - \xi]^T[\Lambda^T(X^T X)^+\Lambda]^{-1}[\Lambda^T u^* - \xi]$$

$$u^* = (X^T X)^+ X^T Y \quad .$$

Theorem 1.8 [21] — Let $Y \sim N(u,V)$. Let $S_1 = Y^T AY$ and $S_2 = Y^T BY$. Then we have

(i) $E[S_1] = \text{Trace}[AV] + u^T Au$

(ii) $V[S_1] = 2\ \text{Trace}[AV]^2 + 4u^T AVAu$

(iii) $\text{Cov}\ [S_1,S_2] = 2\ \text{Trace}[AVBV] + 4u^T AVBu$

Theorem 1.9 [15] — Let $Y \sim N(u,V)$. Let $T = AY$, $S_1 = Y^T BY$ and $S_2 = Y^T FY$, then

(i) T is independent of S_1 if and only if $AVB = 0$

(ii) S_1 and S_2 are independent if and only if $BVF = 0$.

Theorem 1.10 [15] — Let A be an n×n symmetric matrix with rank n . There exists a nonsingular matrix P such that $P^T AP = I$.

Theorem 1.11 [15] – Let A be an $n \times n$ symmetric matrix with rank n. There is an orthogonal matrix H such that $H^T AH = D$ where D is a diagonal matrix of eigenvalues.

Theorem 1.12 [30] – Let $Y \sim N(u,V)$. Then $\mathrm{Cov}\,(Y, Y^T AY) = 2VAu$.

Theorem 1.13 [27] – Let $t_1, \cdots, t_n$ be unbiased estimates of θ, and let V denote the covariance matrix of the t_i's. If we want to choose $a^T = (a_1 \cdots, a_n)$ so that

$$E\left[\sum_{i=1}^{n} a_i t_i\right] = \theta$$

and

$$V\left[\sum_{i=1}^{n} a_i t_i\right]$$

is a minimum, the optimum choice of a is

$$a = \frac{V^{-1}\ell}{\ell^T V^{-1}\ell}$$

where $\ell^T = (1, \cdots, 1)$.

Theorem 1.14 [27] — Let $Y = W\beta + e$ where W is an $n \times p$ matrix of rank $q < p$, β is a constant vector and $e \sim N(0, \sigma^2 I)$.

then

(1) minimum variance unbiased estimate of $\lambda^T\beta$ is $\lambda^T\hat{\beta}$, where $\hat{\beta} = (W^TW)^+W^TY = W^TY$

(2) minimum variance unbiased estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-q}[Y - W\hat{\beta}]^T[Y - W\hat{\beta}]$$

$$= \min_{\beta} \frac{[Y - W\beta]^T[Y - W\beta]}{n-q}$$

Theorem 1.15 [15] — Let A, B, C be $n \times n$ matrices, then

(i) $\mathrm{Trace}(AB) = \mathrm{Trace}(BA)$

(ii) $\mathrm{Trace}(ABC) = \mathrm{Trace}(BCA) = \mathrm{Trace}(CAB)$

(iii) If A is also idempotent

$\mathrm{rank}(A) = \mathrm{Trace}(A)$.

Theorem 1.A — Let $A^2 = A$ and $B^2 = B$. If $A-B$ is positive semidefinite, then $AB = B$.

Proof — Let $B = [b_1, \cdots, b_r]$. In order to show $AB = B$, we will show that $Ab_i = b_i$ for arbitrary i. Now suppose $b_i = 0$, then $Ab_i = b_i$. Suppose $b_i \neq 0$

then $b_i^T[A - B]b_i \geq 0$. Hence $b_i^T[A]b_i \geq b_i^T Bb_i$. But B is idempotent. Hence $Bb_i = b_i$. Therefore $b_i^T Ab_i \geq b_i^T b_i$. Let $S = \frac{b_i^T Ab_i}{b_i^T b_i}$. Now $S \geq 1$. But by Theorem 1.5, we see that $\max \frac{x^T Ax}{x^T x}$ occurs at the maximum eigenvalue of A . Hence $S \leq \lambda_{MAX}$. But A is idempotent. Hence $\lambda_{MAX} = 1$. Therefore $S = 1$. Hence $b_i^T Ab_i = b_i^T b_i$ or $b_i^T[I - A]b_i = 0$. Now since A is idempotent $I-A$ is idempotent. Hence $b_i^T[I - A]b_i = b_i^T[I - A][I - A]b_i$. Let $z_i = [I - A]b_i$. Now $0 = b_i^T[I - A]b_i = z_i^T z_i$. Hence $z_i = 0$ or $Ab_i = b_i$. Hence $AB = B$.

<u>Theorem 1.B</u> — Let $R^2 = \min_z [T - Az]^T[T - Az]$. Then $R^2 = T^T[I - AA^+]T$.

<u>Proof</u> — Now it is well known that the z that minimizes $[T - Az]^T[T - Az]$ is a z such that $A^T Az = A^T T$. Now it is also well known that the normal equations always have a solution, hence by Theorem 1.1, one solution is $z = (A^T A)^+ A^T T = A^+ T$. Hence

$$\begin{aligned} R^2 &= [T - AA^+T]^T[T - AA^+T] \\ &= [[I - AA^+]T]^T[I - AA^+]T \\ &= T^T[I - AA^+][I - AA^+]T \\ &= T^T[I - AA^+]T \end{aligned}$$

Theorem 1.C — Let $R^2 = \min\, [Y - Xu]^T[Y - Xu]$ subject to $\Lambda^T u = \xi$ where Λ^T is of full row rank, then $R^2 = [Y - X\bar{u}]^T[Y - X\bar{u}]$ where

$$\bar{u} = [I - (X^TX)^{-1}\Lambda[\Lambda^T(X^TX)^{-1}\Lambda]^{-1}\Lambda^T](X^TX)^{-1}X^TY$$
$$+ (X^TX)^{-1}\Lambda[\Lambda^T(X^TX)^{-1}\Lambda]^{-1}\xi \quad .$$

Proof — Let us construct the Lagrangian function

$$L(u,\rho) = [Y - Xu]^T[Y - Xu] + 2\rho^T[\Lambda^T u - \xi]$$

Now

$$\frac{\partial L(u,\rho)}{\partial u} = -2X^TY + 2(X^TX)u + 2\Lambda\rho$$

and

$$\frac{\partial L(u,\rho)}{\partial \rho} = 2(\Lambda^T u - \xi)$$

Setting the partials equal to zero, we have

$$(X^TX)\bar{u} = X^TY - \Lambda\bar{\rho}$$

$$\Lambda^T\bar{u} = \xi$$

hence

$$\bar{u} = (X^TX)^{-1}X^TY - (X^TX)^{-1}\Lambda\bar{\rho}$$

Now, since $\Lambda^T\bar{u} = \xi$, we have

$$\xi = \Lambda^T(X^TX)^{-1}X^TY - \Lambda^T(X^TX)^{-1}\Lambda\bar{\rho}$$

or

$$\bar{\rho} = -(\Lambda^T(X^TX)^T\Lambda)^{-1}\xi + (\Lambda^T(X^TX)^{-1}\Lambda)^{-1}\Lambda^T(X^TX)^{-1}X^TY \ .$$

Hence

$$\bar{u} = (X^TX)^{-1}X^TY + (X^TX)^{-1}\Lambda[\Lambda^T(X^TX)^{-1}\Lambda]^{-1}\xi$$

$$- (X^TX)^{-1}\Lambda(\Lambda^T(X^TX)^{-1}\Lambda)^{-1}\Lambda^T(X^TX)^{-1}X^TY$$

$$= [I - (X^TX)^{-1}\Lambda(\Lambda^T(X^TX)^{-1}\Lambda)^{-1}\Lambda^T](X^TX)^{-1}X^TY$$

$$+ (X^TX)^{-1}\Lambda(\Lambda^T(X^TX)^{-1}\Lambda)^{-1}\xi \ .$$

Hence $R^2 = [Y - X\bar{u}]^T[Y - X\bar{u}]$.

<u>Theorem 1.D</u> — Let $R_0^2 = \min_{u\varepsilon\omega_1} [Y - Xu]^T[Y - Xu]$ and $R_1^2 = \min_{u\varepsilon\omega_2} [Y - Xu]^T[Y - Xu]$. If $\omega_2 \subset \omega_1$, then $R_0^2 \leq R_1^2$.

<u>Proof</u> — Let $\bar{u}\varepsilon\omega_2$ such that $R_1^2 = [Y - X\bar{u}]^T[Y - X\bar{u}]$. Let $\tilde{u}\varepsilon\omega_1$ such that $R_0^2 = [Y - X\tilde{u}]^T[Y - X\tilde{u}]$. Now $\bar{u}\varepsilon\omega_2$ implies $\bar{u}\varepsilon\omega_1$. Hence

$$R_0^2 = [Y - X\tilde{u}]^T[Y - X\tilde{u}] \leq [Y - X\bar{u}]^T[Y - X\bar{u}]$$

$$= R_1^2 \ .$$

Theorem 1.E — Let $R^2 = \min\,[Y - Xu]^T[Y - Xu]$ subject to $\Lambda^T u = \xi$, where Λ^T is r×p of rank r, then $\frac{R^2}{\sigma^2} \sim \chi^2(k,\lambda)$, where $k = n - p + r$ and λ is zero if $\Lambda^T u = \xi$.

Proof — Now R^2 is the min $[Y-Xu]^T[Y-Xu]$ subject to $u \varepsilon \omega = \{u/\Lambda^T u = \xi\}$. By Theorem 1.1, all solutions to $\Lambda^T u = \xi$ can be written as $u = \xi_0 + Bz$, where $\Lambda^T \delta_0 = \xi$, $\Lambda^T B = 0$ and z is arbitrary. Hence $\omega = \{u/u = \xi_0 + Bz\}$. Therefore

$$R^2 = \min_z\,[Y - X\xi_0 - XBz]^T[Y - X\xi_0 - (XB)z] \quad .$$

Let $T = Y - X\xi_0$ and $A = XB$

$$R^2 = \min_z\,[T - Az]^T[T - Az] \quad .$$

This is in the form of the classic least-squares problem. Hence by Theorem 1.B, $R^2 = T^T[I - AA^+]T$. Now by Theorem 1.2, $T \sim N[X(u - \xi_0), \sigma^2 I]$. Since $I-AA^+$ is idempotent and by Theorem 1.4, $\frac{T^T[I - AA^+]T}{\sigma^2}$ is $\chi^2(k,\lambda)$, where $k = \text{rank}[I - AA^+]$ and

$$\lambda = [X(u - \xi_0)]^T[I - AA^+]X(u - \xi_0)] \quad .$$

Now by Theorems 1.1 and 1.15,

$$\begin{aligned} k &= \mathrm{rank}[I - AA^{+}] = \mathrm{Trace}[I - AA^{+}] \\ &= n - \mathrm{Trace}[AA^{+}] = n - \mathrm{rank}(A) \\ &= n - \mathrm{rank}(XB) = n - \mathrm{rank}(B) \\ &= n - (p - r) = n - p + r \ . \end{aligned}$$

Now, if $\Lambda^T u = \xi$, then $u = \xi_0 + Bz$, hence $u - \xi_0 = Bz$. Therefore

$$\begin{aligned} \lambda &= [XBz]^T[I - (XB)(XB)^{+}]XBz \\ &= (XBz)^T[XB - XB(XB)^{+}XB]z \\ &= 0 \ . \end{aligned}$$

Therefore if $\Lambda^T u = \xi$, then $\lambda = 0$ and $R^2 \sim \chi^2(k)$.

Theorem 1.F — Let $Q = Q_1 + Q_2$ where $Q \sim \chi^2(k,\lambda)$ and $Q_1 \sim \chi^2(k_1,\lambda_1)$. If $Q_2 = Q - Q_1$ is nonnegative, then $Q_2 \sim \chi^2(k - k_1, \lambda - \lambda_1)$ and Q_1 and Q_2 are independent.

Proof — We can assume without loss of generality that $Q = \dfrac{Y^T AY}{\sigma^2}$ and $Q_1 = \dfrac{Y^T BY}{\sigma^2}$ hence $Q_2 = \dfrac{Y^T[A - B]Y}{\sigma^2}$, where $Y \sim N(u,\sigma^2 I)$. Now A-B is positive semidefinite and by Theorem 1.4, $A^2 = A$ and $B^2 = B$. Hence by

Theorem 1.A, $AB = B$. Therefore let us consider $I = [I - A] + [A - B] + B$; or

$$Y^TY = Y^T[I - A]Y + Y^T[A - B]Y + Y^TBY .$$ Now

(i) $[I - A][A - B] = A - A - B + B = 0$

(ii) $[A - B]B = B - B = 0$

(iii) $[I - A]B = B - B = 0$

hence by Theorem 1.3, we have

(i) $\dfrac{Y^T[I - A]Y}{\sigma^2} \sim \chi^2(n - k, \lambda_0)$

(ii) $\dfrac{Y^T[A - B]Y}{\sigma^2} \sim \chi^2(k_2, \lambda_2)$

(iii) $\dfrac{Y^T[B]Y}{\sigma^2} \sim \chi^2(k_1, \lambda_1)$

(iv) $\dfrac{Y^T[I - A]Y}{\sigma^2}$, $\dfrac{Y^T[A - B]Y}{\sigma^2}$ and $\dfrac{Y^TBY}{\sigma^2}$ are independent.

Now $u^Tu = \lambda_0 + \lambda_2 + \lambda_1$ but $\lambda_0 = u^Tu - \lambda$, hence $\lambda_2 = \lambda - \lambda_1$, and $n = n - k + k_2 + k_1$ hence $k_2 = k - k_1$.

CHAPTER 2. THE MODEL

2.0 Preliminaries

In this chapter we will assume that we have sampled p univariate populations, where each population has a mean and a common variance. Also, let us suppose there may be certain restrictions known about the means. Now, while each population may have been sampled a different number of times, only those populations from which at least one sample was taken shall be included in the model. These observations can be expressed as

$$y_{ijk} = u_{ij} + e_{ijk}$$

where the means satisfy the relations

$$\theta_\ell^T u = \xi_\ell \quad (\ell = 1 \cdots r)$$

where y_{ijk} is the k*th* observation from the (ij)*th* population:

u is the vector of u_{ij} in some order,

u_{ij} is the mean of (ij)*th* population,

e_{ijk} are uncorrelated random variables such that $E(e_{ijk}) = 0$, $V(e_{ijk}) = \sigma^2$,

θ_ℓ^T is the ℓth restriction on the u_{ij}'s, and

ξ_ℓ is known.

Naturally, the number of subscripts is essentially unlimited; however, to simplify the notation, the number of subscripts will be kept to three. In order to better convey the concepts in this section, we will consider the following example.

Suppose we had six populations indexed by two subscripts (ij), that is, (11, 12, 21, 22, 31, 32). Also, suppose we took the following number of observations from each population; 2 from population 1, 2 from population 2, 5 from population 3, 3 from population 4, 1 from population 5, and 5 from population 6. Let us further suppose that we know

$$u_{11} - u_{12} - u_{21} + u_{22} = 0$$

$$u_{11} - u_{12} - u_{31} + u_{32} = 0$$

where u_{ij} is the mean of the $(ij)th$ population.
We can express the above as

$$y_{ijk} = u_{ij} + e_{ijk}$$

subject to $u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0$ for all i, i', j, j' . $i = 1,2,3$; $j = 1,2$; $k = 1,\cdots n_{ij}$.

In order to simplify the presentation, we can rewrite the general model in matrix notation. We have

$$Y = Xu + e \qquad (2.0)$$

subject to $\theta^T u = \xi$

where

Y is the $n\times 1$ vector of observations,

u is the $p\times 1$ vector of cell means,

X is the $n\times p$ design matrix,

e is a random variable such that $E[e] = 0$, $E[ee^T] = \sigma^2 I$,

θ^T is an $r\times p$ matrix of rank r that represents the restrictions known about the means,

ξ is an $r\times 1$ known vector.

Let us express our example in matrix notation. We have

$$Y = Xu + e$$

subject to $\theta^T u = 0$

where

Y is an 18×1 vector

X is an 18×6 matrix

u is a 6×1 vector

e is an 18×1 vector

θ^T is a 2×6 matrix

ξ is a 2×1 vector

and where

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{215} \\ y_{221} \\ y_{222} \\ y_{223} \\ y_{311} \\ y_{321} \\ y_{322} \\ y_{323} \\ y_{324} \\ y_{325} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \\ u_{21} \\ u_{22} \\ u_{31} \\ u_{32} \end{bmatrix} + e$$

subject to $\begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} u = 0$.

Because of the formulation of the model, X^TX will always be diagonal with diagonal elements equal to the number of observations from each population. One statistic that will frequently be encountered is $u^* = (X^TX)^{-1}X^TY$ which is nothing more than the vector of cell means. In our example,

$$(X^TX) = \text{Diag}[2,2,5,3,1,5]$$

and

$$(u^*)^T = [Y_{11\cdot}/2, Y_{12\cdot}/2, Y_{21\cdot}/5, Y_{22\cdot}/3, Y_{31\cdot}/1, Y_{32\cdot}/5] .$$

We shall now consider point estimation and test of hypothesis for the model described by Eq. (2.0). Note that the restriction is an essential part of the model.

2.1 Point Estimation

In this section we want to consider estimation of functions of the u_{ij} . For the moment, we shall consider only linear functions of the u_{ij} . In the classical analysis of experimental design models, the question of estimability is raised, that is, not all linear functions yield unbiased estimates. In particular, Graybill [15] gives the following definition of estimability.

> "...A parameter (or a function of the parameters) is said to be linearly estimable if there exists a linear combination of the observations whose expected value is equal to the parameter (or the function of the parameters)."

Now in the model described by Eq. (2.0), we see that u_{ij} is always estimable. One obvious estimate is y_{ij1} . Hence, any linear combination of the u_{ij} is estimable. A better estimate of u_{ij} would be the mean of the observations of the $(i,j)th$ population, or in other words, the cell mean. Hence, one estimate of $\lambda^T u$ would be $\lambda^T u^* = \lambda^T (X^T X)^{-1} X^T Y$. Note that $E[\lambda^T u^*] = \lambda^T [X^T X]^{-1} X^T X u = \lambda^T u$. Therefore, for the model defined by Eq. (2.0), there is no need to be concerned with the question of estimability. Now we shall consider the problem of finding the best (minimum variance) linear unbiased estimate of $\lambda^T u$.

Theorem 2.1 — Suppose we wish to estimate $\lambda^T u$. The Best Linear Unbiased Estimate (B.L.U.E.) for $\lambda^T u$ is

$\lambda^T\hat{u}$ where $\hat{u}$ is that value of u that minimizes $[Y - Xu]^T[Y - Xu]$ subject to $\theta^T u = \xi$.

<u>Proof</u> — The proof will proceed in the following manner: (1) we shall find $\hat{u}$, (2) we shall find the B.L.U.E. of $\lambda^T u$ and, (3) then observe that the B.L.U.E. of $\lambda^T u$ is equal to $\lambda^T\hat{u}$.

Let us find $\hat{u}$. In order to do this, we shall construct the Lagrangian function

$$L(u,\delta) = [Y - Xu]^T[Y - Xu] + 2\delta^T[\theta^T u - \xi]$$

Now

$$\frac{\partial L(u,\delta)}{\partial u} = -2X^T Y + 2(X^T X)u + 2\theta\delta$$

and

$$\frac{\partial L(u,\delta)}{\partial \delta} = \theta^T u - \xi$$

Upon setting the partials equal to zero, we obtain

$$(X^T X)\hat{u} + \theta\hat{\delta} = X^T Y$$

$$\theta^T\hat{u} = \xi$$

Now

$$\hat{u} = (X^TX)^{-1}X^TY - (X^TX)^{-1}\theta\hat{\delta}$$

But

$$\theta^T\hat{u} = \xi$$

hence

$$\hat{\delta} = [\theta^T(X^TX)^{-1}\theta]^{-1}[\theta^T(X^TX)^{-1}X^TY]$$

$$- [\theta^T(X^TX)^{-1}\theta]^{-1}\xi$$

therefore

$$\hat{u} = [I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]u^*$$

$$+ (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\xi \tag{2.1}$$

or

$$\hat{u} = u^* - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}[\theta^Tu^* - \xi] \tag{2.2}$$

where $u^* = (X^TX)^{-1}X^TY$.

Let us now find the B.L.U.E. of λ^Tu. The problem can be restated as one of finding $a^TY + d$, such that $V[a^TY + d] = \sigma^2a^Ta$ is a minimum subject to $E[a^TY + d] = \lambda^Tu$ given $\theta^Tu = \xi$. Let us consider d as $d = \gamma^T\xi$. Now $E[a^TY + d] = a^TXu + \gamma^T\xi = \lambda^Tu$

whenever $\theta^T u = \xi$ or $a^T X u + \gamma^T \theta^T u = \lambda^T u$, which means $a^T X + \gamma^T \theta^T = \lambda^T$. Hence, we want to minimize $\sigma^2 a^T a$ subject to $a^T X + \gamma^T \theta^T = \lambda^T$. The Lagrangian function is

$$L(a,\gamma,\rho) = \sigma^2 a^T a + 2\sigma^2[\lambda^T - a^T X - \gamma^T \theta^T]\rho$$

Taking partials, we get

$$\frac{\partial L(a,\gamma,\rho)}{\partial a} = 2\sigma^2 a - 2\sigma^2 X\rho$$

$$\frac{\partial L(a,\gamma,\rho)}{\partial \gamma} = 2\theta^T \rho$$

$$\frac{\partial L(a,\gamma,\rho)}{\partial \rho} = \lambda^T - \hat{a}^T X - \gamma^T \theta^T$$

Setting the partials equal to zero, we have

$$X\hat{\rho} = \hat{a}$$

$$\theta^T \hat{\rho} = 0$$

$$X^T \hat{a} + \theta\hat{\gamma} = \lambda$$

or $X^T X \hat{\rho} = X^T \hat{a}$, which means

$$\hat{\rho} = (X^T X)^{-1} X^T \hat{a} = (X^T X)^{-1}[\lambda - \theta\hat{\gamma}]$$

But $\theta^T\hat{\rho} = 0$, hence

$$\theta^T(X^TX)^{-1}[\lambda - \theta\hat{\gamma}] = 0$$

or

$$[\theta^T(X^TX)^{-1}\lambda - \theta^T(X^TX)^{-1}\theta\hat{\gamma}] = 0$$

or

$$\hat{\gamma} = [\theta^T(X^TX)^{-1}\theta]^{-1}\theta^T[X^TX]^{-1}\lambda$$

Since $\hat{a} = X\hat{\rho}$, we have

$$\hat{a} = X(X^TX)^{-1}[\lambda - \theta[\theta^T(X^TX)^{-1}\theta]^{-1}\theta^T(X^TX)^{-1}\lambda]$$

Therefore, the B.L.U.E. of λ^Tu is

$$\begin{aligned}\hat{a}^TY + \hat{d} &= \theta^T[I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T](X^TX)^{-1}X^TY \\ &\quad + \lambda^T(X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta]^{-1}\xi \\ &= \lambda^T\{[I - (X^TX)^{-1}\theta[\theta^T(X^TX)^{-1}\theta]^{-1}\theta^T]u* \\ &\quad + (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\xi\} \qquad (2.3)\end{aligned}$$

But from Eq. (2.1), we see that $\hat{a}^TY + \hat{d}$ is just $\lambda^T\hat{u}$. Hence, the B.L.U.E. of λ^Tu is $\lambda^T\hat{u}$. One immediate consequence of the above is that if $\lambda = I$, then the B.L.U.E. of u is $\hat{u}$ (given $\theta^Tu = \xi$). And if there are no restrictions on the u_{ij} (that is, $\theta^T = 0$), then $\hat{u} = u^*$, a well-known result. The following theorem provides us with an unbiased estimate of σ^2 .

<u>Theorem 2.2</u> — Let $\hat{\sigma}^2 = \frac{1}{n - p + r}[Y - X\hat{u}]^T[Y - X\hat{u}]$ then $E[\hat{\sigma}^2] = \sigma^2$.

<u>Proof</u> — Let us consider

$$Y - X\hat{u} = Y - X\{[I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T) \cdot (X^TX)^{-1}X^TY] + (X^TX)^{-1}(\theta^T(X^TX)^{-1}\theta)^{-1}\xi\}$$

Let

$$A = I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T$$

then

$$\begin{aligned}
Y - X\hat{u} &= Y - XA(X^TX)^{-1}X^T[Xu + e] \\
&\quad - X(X^TX)^{-1}(\theta^T(X^TX)^{-1}\theta)^{-1}\xi \\
&= Y - XAu - XAe - X(X^TX)^{-1}(\theta^T(X^TX)^{-1}\theta)^{-1}\xi \\
&= X[I - A]u + [I - XA(X^TX)^{-1}X^T]e \\
&\quad - X(X^TX)^{-1}(\theta^T(X^TX)^{-1}\theta)^{-1}\xi \\
&= X(X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}[\theta^Tu - \xi] \\
&\quad + [I - XA(X^TX)^{-1}X^T]e
\end{aligned}$$

But $\theta^T u = \xi$, hence

$$Y - X\hat{u} = [I - XA(X^TX)^{-1}X^T]e \qquad (2.4)$$

We are now going to show that $I - XA(X^TX)^{-1}X^T$ is idempotent. We will first show that A is idempotent. Let us set $Z = (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T$. Then,

$$[(X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T][(X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]$$

$$= (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta^{-1}\theta^T$$

Hence Z is idempotent; therefore, $A = I - Z$ is idempotent. Now

$$XA(X^TX)^{-1}X^TXA(X^TX)^{-1}X^T = XAA(X^TX)^{-1}X^T$$

$$= XA(X^TX)^{-1}X^T$$

Therefore, $I - XA(X^TX)^{-1}X^T$ is idempotent and

$$[I - XA(X^TX)^{-1}X^T]^T = I - X(X^TX)^{-1}A^TX^T]$$

$$= I - X(X^TX)^{-1} \cdot [I -$$

$$\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T(X^TX)^{-1}]X^T$$

$$= I - X[I - (X^TX)^{-1}$$

$$\cdot \theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T(X^TX)^{-1}]X^T$$

$$= I - XA(X^TX)^{-1}X^T$$

Let us set $W = I - XA(X^TX)^{-1}X^T$. Hence, we see that from the above, W is symmetric and idempotent. Hence,

$$[Y - X\hat{u}]^T[Y - X\hat{u}] = e^TW^TWe$$

$$= e^TWe \qquad (2.5)$$

Now by Theorem 1.15,

$$E[e^TWe] = \sigma^2\text{Trace}[W]$$

$$= \sigma^2\text{Trace}[I - XA(X^TX)^{-1}X^T]$$

$$= \sigma^2[n - \text{Trace}(XA(X^TX)^{-1}X^T)]$$

$$= \sigma^2[n - \text{Trace}A]$$

$$= \sigma^2[n - \text{Trace}[I$$

$$- (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]]$$

$$= \sigma^2[n - [p$$

$$- \text{Trace}[(X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]]]$$

$$= \sigma^2[n - [p - \mathrm{Trace}[I_{r\times r}]]$$

$$= \sigma^2[n - p + r]$$

Hence,

$$\hat{\sigma}^2 = \frac{1}{n - p + r}[Y - X\hat{u}]^T[Y - X\hat{u}]$$

is an unbiased estimate of σ^2 .

Up to this point, we have made no assumptions about the distribution of the e vector. We shall now assume that $e \sim N(0,\sigma^2 I)$. With this assumption, we are able to prove the following theorem.

<u>Theorem 2.3</u> — Suppose we are given

$$Y = Xu + e \quad \text{subject to} \quad \theta^T u = \xi$$

where $e \sim N(0,\sigma^2 I)$ and where Y, X, u, θ^T, ξ are defined by equation (2.0). The estimates, $\hat{u}$ and $\hat{\sigma}^2$, of u and σ^2 , where

$$\hat{u} = [I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T][(X^TX)^{-1}X^TY] + (X^TX)^{-1}\theta[\theta^T(X^TX)^{-1}\theta]^{-1}\xi$$

$$\hat{\sigma}^2 = \frac{1}{n - p + r}[Y - X\hat{u}]^T[Y - X\hat{u}]$$

have the following properties:

(a) Unbiased

(b) Consistent

(c) Efficient

(d) Sufficient

(e) Complete

(f) Minimum variance

(g) $\hat{u}$ ~ Multivariate Singular Normal

(h) $\left(\frac{n - p + r}{\sigma^2}\right) \hat{\sigma}^2 \sim X^2(n - p + r)$

(i) $\hat{u}$ and $\hat{\sigma}^2$ are independent

Proof — The likelihood function is

$$f(e;u,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \text{EXP} \left\{ - \frac{1}{2\sigma^2} [Y - Xu]^T [Y - Xu] \right\} \tag{2.6}$$

$$\text{subject to} \quad \theta^T u = \xi$$

In order to establish some of the properties, we will show that $\hat{u}$ and $\hat{\sigma}^2$ are essentially maximum likelihood estimates. We want the values of u and σ^2 that maximizes Eq. (2.6) where u must satisfy $\theta^T u = \xi$. However, we shall maximize ln $[f(e,u,\sigma^2)]$

instead of $f(e,u,\sigma^2)$. Therefore, we want to maximize

$$-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}[Y - Xu]^T[Y - Xu]$$

subject to $\theta^T u = \xi$.

Constructing the Lagrangian function, we get

$$L(u,\sigma^2,\delta) = \ln[f(e,u,\sigma^2)] + \delta^T[\theta^T u - \xi]$$

Taking the partials of $L(u,\sigma^2,\delta)$ with respect to u, σ^2, δ , we obtain

$$\frac{\partial L}{\partial u} = \frac{1}{\sigma^2}[X^TY - (X^TX)u] + \theta\delta$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{[Y - Xu]^T[Y - Xu]}{2\sigma^4}$$

$$\frac{\partial L}{\partial \delta} = \theta^T u - \xi$$

Equating the partials to zero, we have

$$(X^TX)\hat{u} - \theta\hat{\delta}\tilde{\sigma}^2 = X^TY$$

$$\tilde{\sigma}^2 = \frac{1}{n}[Y - X\hat{u}]^T[Y - X\hat{u}]$$

$$\theta^T\hat{u} = \xi$$

Now, solving for $\hat{u}$ in the first equation we have

$$\hat{u} = (X^TX)^{-1}X^TY + (X^TX)^{-1}\theta\hat{\delta}\tilde{\sigma}^2$$

But since $\theta^T\hat{u} = \xi$, we have

$$\xi = \theta^T(X^TX)^{-1}X^TY + \theta^T(X^TX)^{-1}\theta\hat{\delta}\tilde{\sigma}^2$$

or

$$\hat{\delta}\tilde{\sigma}^2 = -[\theta^T(X^TX)^{-1}\theta]^{-1}[\theta^Tu^* - \xi]$$

where

$$u^* = (X^TX)^{-1}X^TY$$

Therefore,

$$\begin{aligned}\hat{u} &= u^* - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}[\theta^Tu^* - \xi] \\ &= [I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]u^* \\ &\quad + (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\xi\end{aligned}$$

which is identical to Eq. (2.1). Hence, we know that $\hat{u}$ is (1) unbiased and (2) has minimum variance among estimates which are linear in the observations. It will be shown that owing to sufficiency and completeness, we can extend the class in which $\hat{u}$ is best to include all unbiased estimates of u . Hence, $\hat{u}$ and $\tilde{\sigma}^2$ are maximum likelihood estimates. They are consistent and asymptotically efficient as all the n_{ij}'s increase. Hence, $\hat{\sigma}^2$ is also consistent and efficient; and by Theorem 2.2, $\hat{\sigma}^2$ is also unbiased.

Now let us consider the following

$$[Y - Xu]^T[Y - Xu] = [\hat{u} - u]^T(X^TX)[\hat{u} - u] + [Y - X\hat{u}]^T[Y - X\hat{u}] \qquad (2.7)$$

given $\theta^T u = \xi$.

The proof of Eq. (2.7) is as follows

$$[Y - X\hat{u}]^T[Y - X\hat{u}] + (\hat{u} - u)^T(X^TX)(\hat{u} - u)$$
$$= e^T[W]e + \hat{u}^T(X^TX)\hat{u} - 2u^T(X^TX)\hat{u} + u^T(X^TX)u$$

But $X\hat{u} = Xu + (I - W)e$ from Eq. (2.4). Therefore, we have

$$\text{(i)} \quad \hat{u}^T(X^TX)\hat{u} = u^TX^TXu + 2u^TX^T(I - W)e + e^T[I - W^T][I - W]e$$

$$\text{(ii)} \quad -2u^TX^TX\hat{u} = -2u^TX^TXu - 2u^TX^T(I - W)e$$

Therefore

$$(Y - X\hat{u})^T(Y - X\hat{u}) + (\hat{u} - u)^T(X^TX)(\hat{u} - u)$$
$$= e^T[W]e + e^T[I - W^T][I - W]e$$
$$= e^TWe + e^Te - e^TW^Te - e^TWe + e^TW^TWe$$

$$= e^T e$$

$$= [Y - Xu]^T [Y - Xu]$$

Thus, Eq. (2.7) is an identity. Therefore, the likelihood equation can now be written as

$$f(e;u,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \text{EXP} \left\{ -\frac{1}{2\sigma^2} [(n - p + r)\hat{\sigma}^2 + (\hat{u} - u)^T (X^T X)(\hat{u} - u)] \right\}$$

Hence, $\hat{u}$ and $\hat{\sigma}^2$ are jointly sufficient for u and σ^2. It can also be shown that $\hat{u}$ and $\hat{\sigma}^2$ are complete (Ref. [15]). Since $\hat{u}$ and $\hat{\sigma}^2$ are complete, sufficient statistics, if a function can be found such that $E[f(\hat{u},\hat{\sigma}^2)] = g(u,\sigma^2)$, then f is the minimum variance, unbiased estimate of $g(u,\sigma^2)$. Hence, we see that both $\hat{u}$ and $\hat{\sigma}^2$ are unbiased minimum variance estimates in the class of unbiased estimates of u and σ^2.

Let us now find the distribution of $\hat{u}$. From Eq. (2.1), we have

$$\hat{u} = Au^* + (X^T X)^{-1}\theta(\theta^T (X^T X)^{-1}\theta)^{-1}\xi$$

$$= A(X^T X)^{-1} X^T Y + (X^T X)^{-1}\theta(\theta^T (X^T X)^{-1}\theta)^{-1}\xi \tag{2.8}$$

Hence

$$
\begin{aligned}
V[\hat{u}] &= \sigma^2(A(X^TX)^{-1}X^TX(X^TX)^{-1}A^T) \\
&= \sigma^2 A(X^TX)^{-1}A^T \\
&= \sigma^2 A(X^TX)^{-1}[I - \theta^T(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T(X^TX)^{-1}] \\
&= \sigma^2 A[I - \theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T](X^TX)^{-1} \\
&= \sigma^2 AA(X^TX)^{-1} \\
&= \sigma^2 A(X^TX)^{-1}
\end{aligned}
$$

and $\hat{u}$ is normal by Theorem 1.2. Note that since A is idempotent, $V[\hat{u}]$ is singular and hence u has a multivariate singular normal distribution. The reason for this is that u is constrained to lie in a p-r dimensional subspace of the p space. The implications of this will be taken up when we consider interval estimation.

Let us now consider the distribution of $\hat{\sigma}^2$. Now by Eq. (2.5) we have $(n - p + r)\hat{\sigma}^2 = e^T[W]e$. Therefore, by Theorem 1.4,

$$\frac{e^T(W)e}{\sigma^2} \sim \chi^2(n - p + r)$$

Hence,

$$\frac{n - p + r}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n - p + r)$$

We will now show that $\hat{u}$ and $\hat{\sigma}^2$ are independent. To do this, we observe that from Eq. (2.8) and from the fact that $\theta^T u = \xi$ that

$$\begin{aligned} \hat{u} &= A(X^TX)^{-1}X^T[Xu + e] + (X^TX)^{-1}\theta[\theta^T(X^TX)^{-1}\theta)^{-1}\xi \\ &= Au + (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta^{-1})\xi + A(X^TX)^{-1}X^Te \\ &= u + A(X^TX)^{-1}X^Te \end{aligned} \tag{2.9}$$

Thus $\hat{u} - u = A(X^TX)^{-1}X^Te$. Now, by Theorem 1.9, $\hat{u}$-u and $\hat{\sigma}^2$ are independent if $(A(X^TX)^{-1}X^T)(W) = 0$. But

$$\begin{aligned} A(X^TX)^{-1}X^TW &= A(X^TX)^{-1}X^T[I - XA(X^TX)^{-1}X^T] \\ &= A(X^TX)^{-1}X^T - A(X^TX)^{-1}X^T = 0 \end{aligned} \tag{2.10}$$

Therefore $\hat{u}$-u and $\hat{\sigma}^2$ are independent, which also implies that $\hat{u}$ and $\hat{\sigma}^2$ are independent.

2.2 Interval Estimation

Since the distribution of $\frac{n-p+r}{\sigma^2}\hat{\sigma}^2$ is known to be $\chi^2(n-p+r)$, the procedures for interval estimation of σ^2 are well known and will not be discussed here.

However, since $\hat{u}$ has singular distribution, we shall discuss the problems associated with finding confidence intervals for $\lambda^T u$. For the moment, we will assume that λ^T is a $1\times p$ vector. We will postpone the discussion of simultaneous confidence intervals until Section 2.4.

Suppose we wish to place a confidence interval on $\lambda^T u$. Then, the B.L.U.E. of $\lambda^T u$ is $\lambda^T\hat{u}$ and $V[\lambda^T\hat{u}] = \sigma^2\lambda^T A(X^TX)^{-1}A^T\lambda$. Note that $A(X^TX)^{-1}A^T$ is singular and there is the possibility that $\sigma^2\lambda^T(A(X^TX)^{-1}A^T)\lambda = 0$. Since $(X^TX)^{-1}$ is a positive definite matrix, $\sigma^2\lambda^T(A(X^TX)^{-1}A^T)\lambda = 0$ if and only if $A^T\lambda = 0$. Now we claim $A^T\lambda = 0$ if and only if $\lambda = \theta\gamma$. In other words, the only time $V[\lambda^T\hat{u}] = 0$ is when we are estimating $\lambda^T u = \gamma^T\theta^T u = \gamma^T\xi$, a known constant. The proof of this is: Suppose $A^T\lambda = 0$, then we have by Theorem 1.1,

$\lambda = [I - (A^T)^+ A^T]z$. But $(A^T)^+ = A^T$ since A^T is idempotent. Hence

$$\lambda = [I - A^T A^T]z = [I - A^T]z$$

$$= \theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T(X^TX)^{-1}z$$

$$= \theta\gamma \quad \text{where} \quad \gamma = (\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T(X^TX)^{-1}z$$

Now suppose $\lambda = \theta\gamma$, then

$$A^T\lambda = A^T\theta\gamma = [I - \theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T(X^TX)^{-1}]\theta\gamma$$

$$= [\theta - \theta]\gamma = 0 .$$

Therefore, if u is a nontrivial function of the parameter, we know that $\dfrac{[\lambda^T\hat{u} - \lambda^T u]}{\sqrt{V[\lambda^T\hat{u}]}}$ is distributed as N(0,1). Since $\hat{u} - u = A(X^TX)^{-1}X^Te$, we see that $\lambda^T(\hat{u} - u) = \lambda^T A(X^TX)^{-1}X^Te$ and from Eq. (2.10) $\lambda^T A(X^TX)^{-1}X^TW = \lambda^T 0 = 0$. Hence by Theorem 1.9. $\lambda^T(\hat{u}-u)$ is independent of $\hat{\sigma}^2$. Hence,

$$\frac{\lambda^T\hat{u} - \lambda u}{\sqrt{\sigma^2\lambda^T A(X^TX)^{-1}A^T\lambda}} \Bigg/ \sqrt{\frac{(n - p + r)\hat{\sigma}^2}{\sigma^2}} \quad \text{is a } t \text{ statistic}$$

with n-p+r degrees of freedom. The procedures for finding confidence intervals from this information are well known.

2.3 Hypothesis Testing

In this section we will develop a procedure for testing H_o: $\Lambda^T u = \gamma$ against the two-sided alternative H_A: $\Lambda^T u \neq \gamma$, where Λ^T is an $s \times p$ matrix of rank s. We will use the likelihood ratio test. The following is a brief description of the procedure.

Suppose $x_1, x_2, \cdots, x_n$ are distributed as $L(x,\theta)$, where θ is an element of a set Ω. Now, if we want to test H_o: $\theta \varepsilon \omega$, where ω is a subset of Ω, versus H_A: $\theta \varepsilon [\Omega - \omega]$, we construct

$$\Gamma(x) = \frac{\max\limits_{\theta \varepsilon \omega} L(x,\theta)}{\max\limits_{\theta \varepsilon \Omega} L(x,\theta)}$$

Now we can observe that

(1) $\Gamma(x)$ is a function of x alone and hence a statistic

(2) $0 \leq \Gamma(x) \leq 1$

(3) small values of $\Gamma(x)$ suggest rejection of H_o.

Let us now apply this procedure to our model. Since $Y \sim N[Xu, \sigma^2 I]$ where $\theta^T u = \xi$, we see that

$$L(Y; u, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \text{EXP} \left\{ - \frac{1}{2\sigma^2} [Y - Xu]^T [Y - Xu] \right\}$$

given $\theta^T u = \xi$.

Hence we see that Ω is the set of all u such that $\theta^T u = \xi$ or $\Omega = \{u / \theta^T u = \xi\}$. Now suppose we wish to test H_o: $\Lambda^T u = \gamma$, then ω is the set of all u such that $\Lambda^T u = \gamma$ and $\theta^T u = \xi$; or $\omega = \{u / \Lambda^T u = \gamma$ and $\theta^T u = \xi\}$. Therefore $\Omega - \omega = \{u / \Lambda^T u \neq \gamma$ and $\theta^T u = \xi\}$. And now we can write

$$\Gamma(Y) = \frac{\max\limits_{\begin{bmatrix}\theta^T \\ \Lambda^T\end{bmatrix} u = \begin{bmatrix}\xi \\ \gamma\end{bmatrix}} f(Y; u, \sigma^2)}{\max\limits_{\theta^T u = \xi} f[Y; u, \sigma^2]}$$

Let $B^T = \begin{bmatrix} \theta^T \\ \lambda^T \end{bmatrix}$ and $\delta^T = \begin{bmatrix} \xi \\ \gamma \end{bmatrix}$, then we can write

$$\Gamma(Y) = \frac{\max\limits_{B^T u=\delta} f(Y;u,\sigma^2)}{\max\limits_{\theta^T u=\xi} f(Y;u,\sigma^2)}$$

In Section 2.1, we showed that the values of u and σ^2 that maximized $f(Y;u,\sigma^2)$ subject to $\theta^T u = \xi$ were

$$\hat{u} = [I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]u^* + [X^TX]^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\xi$$

$$\tilde{\sigma}^2 = \frac{1}{n}[Y - X\hat{u}]^T[Y - X\hat{u}] \quad , \quad \text{where}$$

$$u^* = (X^TX)^{-1}X^TY \quad .$$

Therefore

$$\max_{\theta^T u=\xi} f(Y;u,\sigma^2) = (2\pi\tilde{\sigma}^2)^{-\frac{n}{2}} \text{EXP}\left\{-\frac{n}{2}\right\}$$

Likewise it can be shown that the values of u and σ^2 that maximize $f(Y;u,\sigma^2)$ subject to $B^Tu = \delta$ were

$$\bar{u} = [I - (X^TX)^{-1}B[B^T(X^TX)^{-1}B]^{-1}B^T]u^*$$

$$+ (X^TX)^{-1}B(B^T(X^TX)^{-1}B)^{-1}\delta$$

$$\bar{\sigma}^2 = \frac{1}{n}[Y - X\bar{u}]^T[Y - X\bar{u}]$$

Therefore $\sup\limits_{B^Tu=\delta} f(Y;u,\sigma^2) = (2\pi\bar{\sigma}^2)^{-\frac{n}{2}} \, \mathrm{EXP}\left\{-\frac{n}{2}\right\}$.

Hence, we see that

$$\Gamma(Y) = \left(\frac{\bar{\sigma}^2}{\tilde{\sigma}^2}\right)^{-\frac{n}{2}} .$$

The statistic that we will use is $\Gamma(Y)^{-\frac{2}{n}}$. Let us denote $\Gamma(Y)^{-\frac{2}{n}}$ as L^*. Now we reject H_o if $L^* \geq L^*_o$ since $L^* \geq L^*_o$ if and only if $\Gamma_o(Y) \geq \Gamma(Y)$. The problem now is to determine the distribution of L^*. Let us consider the identity

$$L^* = 1 + \frac{\bar{\sigma}^2 - \tilde{\sigma}^2}{\tilde{\sigma}^2} = 1 + L$$

where

$$L = \frac{\bar{\sigma}^2 - \tilde{\sigma}^2}{\tilde{\sigma}^2}$$

Let us note that by Theorem 1.C,

$$n\bar{\sigma}^2 = \min_{u \epsilon \omega_1} [Y - Xu]^T[Y - Xu]$$

and

$$n\tilde{\sigma}^2 = \min_{u \epsilon \omega_2} [Y - Xu]^T[Y - Xu]$$

where

$$\omega_1 = \left\{ u \middle/ \begin{bmatrix} \theta^T \\ \Lambda^T \end{bmatrix} u = \begin{bmatrix} \xi \\ \gamma \end{bmatrix} \right\}$$

and

$$\omega_2 = \{u/\theta^T u = \xi\}$$

Now ω_1 is a subset of ω_2 , hence by Theorem 1.D, $\bar{\sigma}^2$ is greater than or equal to $\tilde{\sigma}^2$. Now by Theorem 1.E, $\frac{n\bar{\sigma}^2}{\sigma^2}$ and $\frac{n\tilde{\sigma}^2}{\sigma^2}$ are $\chi^2(h_1,\lambda_1)$ and $\chi^2(h_2,\lambda_2)$ where $h_1 = n - p + (r + s)$ and λ_1 is zero if $\begin{bmatrix} \theta^T \\ \Lambda^T \end{bmatrix} u = \begin{bmatrix} \xi \\ \gamma \end{bmatrix}$ and where $h_2 = n - p + r$ and $\lambda_2 = 0$ if $\theta^T u = \xi$. But we are given that $\theta^T u = \xi$, hence $\lambda_2 = 0$ and $\lambda_1 = 0$ if $\Lambda^T u = \gamma$. Therefore we have $\frac{n\bar{\sigma}^2}{\sigma^2} = \frac{n\tilde{\sigma}^2}{\sigma^2} + \frac{n}{\sigma^2}(\bar{\sigma}^2 - \tilde{\sigma}^2)$. Now $\frac{n}{\sigma^2}(\bar{\sigma}^2 - \tilde{\sigma}^2) \geq 0$, hence by Theorem 1.F, $\frac{n}{\sigma^2}(\bar{\sigma}^2 - \tilde{\sigma}^2) \sim \chi^2(s,\lambda_1)$ and $\frac{n}{\sigma^2}(\bar{\sigma}^2 - \tilde{\sigma}^2)$ is independent of $\frac{n}{\sigma^2}\tilde{\sigma}^2$. Therefore we see that

$$L = \frac{\dfrac{n(\bar{\sigma}^2 - \tilde{\sigma}^2)}{\sigma^2}}{\dfrac{n\tilde{\sigma}^2}{\sigma^2}}$$

is the ratio of a noncentral χ^2 and a central χ^2. Therefore by Theorem 1.6, $F = \frac{n-p+r}{s} L$ is distributed as a noncentral F with s and $n-p+r$ degrees of freedom and noncentrality parameter λ, where $\lambda = 0$ if H_o: $\Lambda^T u = \gamma$ is true. Now $L^* \geq L_o^*$ if and only if $L \geq L_o$; and $L \geq L_o$ if and only if $F \geq F_o$. Hence we reject H_o: $\Lambda^T u = \gamma$ if $F \geq F_o$. And since we know the distribution of F, it is easy to find F_o.

Let us now consider a computational form for F. By Theorem 1.7, we see that

$$\text{(i)} \quad \min_{B^T u=\delta} [Y - Xu]^T[Y - Xu] = R_o^2 + SS(B^T,\delta)$$

$$\text{(ii)} \quad \min_{\theta^T u=\xi} [Y - Xu]^T[Y - Xu] = R_o^2 + SS(\theta^T,\xi)$$

where

$$R_o^2 = Y^TY - u^{*T}(X^TX)u^*$$

and

$$SS(B^T,\delta) = [B^Tu^* - \delta]^T[B^T(X^TX)^{-1}B]^{-1}[B^Tu^* - \delta]$$

and

$$SS(\theta^T,\xi) = [\theta^Tu^* - \xi]^T[\theta^T(X^TX)^{-1}\theta]^{-1}[\theta^Tu^* - \xi]$$

where $u^* = (X^TX)^{-1}X^TY$. Hence F can be written as

$$F = \frac{n - p + r}{s} \frac{[SS(B^T,\delta)] - SS(\theta^T,\xi)]}{R_o^2 + SS(\theta^T,\xi)} \tag{2.11}$$

Let us also note that

$$\begin{aligned}\lambda &= E[SS(B,\delta) - SS(\theta^T,\xi)] \\ &= [B^Tu - \delta]^T(B^T(X^TX)^{-1}B)^{-1}[B^Tu - \delta] \\ &\quad - (\theta^Tu - \xi)^T[\theta^T(X^TX)^{-1}\theta]^{-1}[\theta^Tu - \xi]\end{aligned}$$

But we are given $\theta^T u = \xi$, hence

$$\lambda = [B^T u - \delta]^T [B^T (X^T X)^{-1} B]^{-1} [B^T u - \delta]$$

$$= \begin{bmatrix} 0 \\ \Lambda^T u - \gamma \end{bmatrix}^T [B^T (X^T X)^{-1} B]^{-1} \begin{bmatrix} 0 \\ \Lambda^T u - \delta \end{bmatrix}$$

$$= [\Lambda^T u - \gamma]^T [B^T (X^T X)^{-1} B]^{-1} [\Lambda^T u - \delta] .$$

Another form for the numerator sum of squares will be given at the end of Section 3.0.

2.3.1 Choice of Λ^T for H_o: $\Lambda^T u = \gamma$.

Some care must be taken when choosing Λ^T for H_o: $\Lambda^T u = \gamma$. For example, suppose we are given the model $Y = Xu + e$ subject to $u_1 - u_2 = 2$. The test H_o: $u_1 - u_2 = 0$ is obviously meaningless. In other words, we must choose Λ^T such that

$$\begin{bmatrix} \theta^T \\ \Lambda^T \end{bmatrix} u = \begin{bmatrix} \xi \\ \gamma \end{bmatrix}$$

is a consistent set of equations and rank $\begin{bmatrix} \theta^T \\ \Lambda^T \end{bmatrix} = r + s$.

2.3.2 Example

Let us consider the example in Section 2.0. There we were given $Y = Xu + e$ subject to $\theta^T u = \xi$, where

$$\theta^T = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \xi = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Now suppose we wanted to test H_o: $u_{1\cdot} = u_{2\cdot} = u_{3\cdot}$ or H_o: $\Lambda_1^T u = 0$ where

$$\Lambda_1^T = \begin{bmatrix} 1 & 1 & -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 \end{bmatrix}$$

Hence we have

$$B_1^T = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & 1 & -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 \end{bmatrix} \qquad \delta = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Hence

$$F = \frac{n - p + r}{s} \, \frac{\left[SS\left(B_1^T, 0\right) - SS(\theta^T, 0)\right]}{SS(\theta^T, 0) + R_o^2}$$

where $n - p + r = 18 - 6 + 2 = 14$ and $s = 2$.

Let us consider another hypothesis. Suppose we wanted to test H_o: $u_{ij} = u_{i'j}$ for all i,i',j. The Λ_2^T matrix would be:

$$\Lambda_2^T = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Now, when θ^T is augmented to Λ_2^T to form $B_2^T = \begin{bmatrix} \theta^T \\ \lambda^T \end{bmatrix}$, we see that B_2^T is a 6×6 matrix of rank 4, that is

$$B_2^T = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

If we add the third row to minus the first row, we obtain the fifth row; and minus the second row plus the fourth row yields the sixth row. Therefore, in light of Section 2.3.1, we would take B_2^T to be

$$B_2^T = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

Let us note that $B_1^T = HB_2^T$ where

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

We will show later that this implies that H_o: $u_{1\cdot} = u_{2\cdot} = u_{3\cdot}$ is equivalent to H_o: $u_{ij} = u_{i'j}$ for i, i', j . This example points out the care needed in selecting the hypothesis and the constructing of the B matrix. By properly setting up the B matrix, the experimenter will know exactly what he is testing and will be guaranteed a valid test.

2.4 Simultaneous Confidence Intervals

In this section, we will be concerned with finding confidence intervals for $\Lambda^T u = \delta$. The procedure is as follows. By Theorem 2.3, we know that $\hat{u} - u \sim N(0, A(X^TX)^{-1}A^T)$. Hence

$$\Lambda^T(\hat{u} - u) \sim N(0, \Lambda^T A(X^TX)^{-1}A^T\Lambda) .$$

Now, from Eq. (2.9) we see that

$$\Lambda^T(\hat{u} - u) = \Lambda^T A(X^TX)^{-1}X^T e .$$

Consider

$$T^2 = \frac{1}{\sigma^2}[\Lambda^T(\hat{u} - u)]^T[\Lambda^T(A(X^TX)^{-1}A^T)\Lambda]^+[\Lambda^T(\hat{u} - u)]$$

or

$$T^2 = \frac{1}{\sigma^2} e^T[X(X^TX)^{-1}A^T\Lambda(\Lambda^TA(X^TX)^{-1}A^T\Lambda)^+\Lambda^TA(X^TX)^{-1}A^T]e$$

$$= e^TG(G^TG)^+G^Te$$

where $G = X(X^TX)^{-1}A^T\Lambda$. But

$$[G(G^TG)^+G^T]G[G^TG]^+G^T = G(G^TG)^+G^T ,$$

hence by Theorem 1.4, $T^2 \sim \chi^2(k)$, where $k = \text{rank}(G)$. Since A is not of full rank, we must consider the possibility that $\Lambda^TA = 0$. However, by following a similar argument as put forth in Section 2.2, we see that $\Lambda^TA = 0$ if and only if $\Lambda^T = \Gamma^T\theta^T$ or $\Lambda^Tu = \Gamma^T\xi$, a known vector. Hence if Λ^Tu is a non-trival function, then $\Lambda^TA \neq 0$. Now

$$T^2 = e^T \frac{[G(G^TG)^+G^T]e}{\sigma^2}$$

is independent of e^TWe since $G(G^TG)^+G^TW = 0$.

Therefore

$$\frac{T^2/k}{\hat{\sigma}^2/\sigma^2} \sim F(h,\ n - p + r) \quad .$$

Hence

$$\Pr\left\{\frac{[\Lambda^T(\hat{u} - u)]^T[\Lambda^T A(X^T X)^{-1}A^T\Lambda]^+[\Lambda^T(\hat{u} - u)]}{k\hat{\sigma}^2} \le F_\alpha\right\} = 1 - \alpha$$

But since $\delta = \Lambda^T u$, the set of δ such that

$$\frac{(\Lambda^T\hat{u} - \delta)^T[\Lambda^T A(X^T X)^{-1}A^T\Lambda]^+[\Lambda^T\hat{u} - \delta]}{k\hat{\sigma}^2} \le F_\alpha$$

is a $(1-\alpha)$ simultaneous interval for $\delta = \Lambda^T u$. Now if $\theta = 0$, then $A = I$ and

$$T^2 = \frac{1}{\sigma^2}[\Lambda^T(u^* - u)]^T[\Lambda^T(X^T X)^{-1}\Lambda]^{-1}[\Lambda^T(u^* - u)]$$

which is the classical result. Numerical examples of the above procedure will be given in Chapter 6.

CHAPTER 3. RELATION TO OTHER MODELS

3.0 Preliminaries

We will now consider the relation between the "u" model discussed in Chapter 2, that is $Y = Xu + e$ subject to $\theta^T u = \xi$, and the classical linear models. The assumptions for the classical models are very similar to the assumptions for the "u" model. For the classical linear models, we assume that we have sampled p univariate populations where each population has a different mean, but all have the same variance.

The functional form for these models is $Y = Xu + e$ subject to $u = P\beta$ or simply

$$Y = W\beta + e \tag{3.1}$$

where $W = XP$ and P is $p \times t$ of rank q. We shall refer to this model as the "β" model.

We shall assume that $e \sim N(0,\sigma^2 I)$, since it is this case that is most interesting. The case in which we assume $E[e] = 0$ and $E[ee^T] = \sigma^2 I$ has limited application and the needed results can be obtained by following a similar line of reasoning as for the case where $e \sim N(0,\sigma^2 I)$. In order to establish a relation between the "u" model and the "β" model, we shall define the concept of statistical equivalence between two models. We say the "u" model and the "β" model are statistically equivalent if:

(i) there is a 1-1 correspondence between linear estimable functions of the parameters, and minimum variance unbiased estimates are identical.

(ii) there is a 1-1 correspondence between testable hypotheses, and the test statistics (under the same criterion) are identical.

Now suppose we are given $Y = W\beta + e$ where $W = XP$, or in other words, $u = P\beta$. Let us consider the following "u" model: $Y = Xu + e$ subject to $\theta^T u = 0$ where:

$$
\begin{aligned}
&(1) \quad \theta^T P = 0 \\
&(2) \quad \mathrm{rank}(\theta^T) = \mathrm{rank}(X) - \mathrm{rank}(XP) = p - q \qquad (3.2) \\
&(3) \quad \theta^T \text{ is } (p-q)\times p \text{ of rank } p-q.
\end{aligned}
$$

In light of the above, we are able to prove the following theorem:

<u>Theorem 3.0</u> – The "u" model as defined by Eq. (3.2) is statistically equivalent to the "β" model as given by Eq. (3.1).

<u>Proof</u> – By Theorem 1.14, the minimum variance unbiased estimate of σ^2 in the "β" model is given by

$$\tilde{\sigma}^2 = \frac{1}{n - q} \min_{\beta} [Y - W\beta]^T[Y - W\beta]$$

where $q = \text{rank}(W)$. And from Theorem 2.3, the minimum variance unbiased estimate of σ^2 in the "u" model is given by

$$\hat{\sigma}^2 = \frac{1}{n - p + r} \min [Y - Xu]^T[Y - Xu]$$

subject to $\theta^T u = 0$. But $r = \text{rank}(\theta^T) = p - q$. Hence $n - p + r = n - q$. Now $\theta^T u = 0$ implies that $u = [I - (\theta^T)^+\theta^T]z$ where z is arbitrary. But we are given that $\theta^T P = 0$; hence $P = [I - (\theta^T)^+\theta^T]G$, where G is chosen so that P is $p \times t$ of rank q.

Let $z = G\gamma$. Hence $u = [I - (\theta^T)^+\theta^T]G\gamma = P\gamma$. Therefore, we have

$$\hat{\sigma}^2 = \frac{1}{n - p + r} \min_{\theta^T u=0} [Y - Xu]^T[Y - Xu]$$

$$= \frac{1}{n - q} \min_{u=P\gamma} [Y - Xu]^T[Y - Xu]$$

$$= \frac{1}{n - p} \min_{\gamma} [Y - W\gamma]^T[Y - W\gamma]$$

$$= \tilde{\sigma}^2 \quad . \tag{3.3}$$

Hence the minimum variance unbiased estimate of σ^2 in the "u" model is identical to the minimum variance unbiased estimate of σ^2 in the "β" model.

Now suppose we wish to estimate $\delta^T\beta$, and that $\delta^T\beta$ is estimable; i.e., there exists an a such that $\delta^T = a^T W$. Since $u = P\beta$, we have

$$\beta = P^+u + [I - P^+P]z \quad .$$

Hence $\delta^T\beta = \delta^T P^+ u + \delta^T[I - P^+P]z$. But $\delta^T = a^T W$ or $\delta^T = a^T XP$. Therefore, we see that

$$\delta^T[I - P^+P]z = a^T XP[I - P^+P]z$$

$$= a^T X[P - P]z = 0$$

Hence $\delta^T\beta = \delta^T P^+ u$. Now let us suppose we want to estimate $\lambda^T u$ in the "u" model. Since $u = P\beta$, we see that $\lambda^T u = \lambda^T P\beta$. Therefore, we have a 1-1 correspondence between estimable functions; that is if we want to estimate $\delta^T\beta$, then we can estimate $\lambda^T u$ where $\lambda^T = \delta^T P^+$, or if we want to estimate $\lambda^T u$ we can estimate $\delta^T\beta$ where $\delta^T = \lambda^T P$.

From Eq. (3.3), we see that

$$\min_{\theta^T u=0} [Y - Xu]^T[Y - Xu] = \min_{\beta} [Y - W\beta]^T[Y - W\beta]$$

Also from Eq. (3.3) we see that $\hat{u}$ is that value of u that minimizes $[Y - Xu]^T[Y - Xu]$ subject to $\theta^T u = 0$; and $\hat{\beta}$ is that value of β that minimizes $[Y - W\beta]^T[Y - W\beta]$.

We will now show that $\hat{u} = P\hat{\beta}$, that is we will show that $P\hat{\beta}$ minimizes $[Y - Xu]^T[Y - Xu]$ subject to

$\theta^T u = 0$. Now let $\bar{u} = P\hat{\beta}$. We see that

$$\theta^T \bar{u} = \theta^T P \hat{\beta} = 0$$

Hence $\bar{u}$ satisfies the constraints. Also

$$\begin{aligned}[Y - X\bar{u}]^T [Y - X\bar{u}] &= [Y - XP\hat{\beta}]^T [Y - XP\hat{\beta}] \\ &= \min_{\beta} [Y - XP\beta]^T [Y - XP\beta] \\ &= \min [Y - Xu]^T [Y - Xu]\end{aligned}$$

subject to $\theta^T u = 0$. Hence $\bar{u} = \hat{u}$, or $\hat{u} = P\hat{\beta}$. Now by Theorem 2.3, the minimum variance unbiased estimate of $\lambda^T u$ is $\lambda^T \hat{u} = \lambda^T P \hat{\beta} = \delta^T \hat{\beta}$ if $\delta^T = \lambda^T P$. But by Theorem 1.14, $\delta^T \hat{\beta}$ is the minimum variance unbiased estimate of $\delta^T \beta$. Hence if we wish to estimate $\delta^T \beta$, then the minimum variance unbiased estimate of $\delta^T \beta$ in the "u" model is $\lambda^T \hat{u}$ where $\lambda^T = \delta^T P^+$. The converse also is true. That is, suppose we wish to estimate $\lambda^T u$, then the minimum variance unbiased estimate of $\lambda^T u$ in the "β" model is $\delta^T \hat{\beta}$ where $\delta^T = \lambda^T P$. Thus condition (i) is satisfied.

If we wish to test H_o: $\delta^T \beta = \gamma$ in the "β" model, we see that this is equivalent to testing H_o: $\lambda^T u = \gamma$ in the "u" model where $\lambda^T P = \delta^T$. The reason for this is the 1-1 correspondence between estimable functions. Now, if we use the likelihood ratio test as a test criterion, the test statistics are identical. This can be established by considering the following:

$$R_1^2 = \min_{\substack{\theta^T u=0 \\ \lambda^T u=\gamma}} [Y - Xu]^T[Y - Xu]$$

$$= \min_{\substack{u=P\beta \\ \lambda^T u=\gamma}} [Y - Xu]^T[Y - Xu]$$

$$= \min_{\delta^T P^+ P\beta=\gamma} [Y - W\beta]\ [Y - W\beta]$$

but $\delta^T = a^T W = a^T XP$. Hence

$$\delta^T P^+ P\beta = a^T XPP^+ P\beta = a^T XP\beta = \delta^T \beta$$

Hence

$$R_1^2 = \min_{\delta^T \beta=\gamma} [Y - W\beta]^T[Y - W\beta]$$

Now the appropriate test in the "u" model is

$$F = \frac{\left[\min\limits_{\substack{\theta^T u=0 \\ \lambda^T u=\gamma}} [Y - Xu]^T[Y - Xu] - \min\limits_{\theta^T u=0} [Y - Xu]^T[Y - Xu]\right]}{s\hat{\sigma}^2},$$

but

$$\min_{\substack{\theta^T u=0 \\ \lambda^T u=\gamma}} [Y - Xu]^T[Y - Xu] = \min_{\delta^T\beta=\gamma} [Y - W\beta]^T[Y - W\beta]$$

and

$$\min_{\theta^T u=0} [Y - Xu]^T[Y - Xu] = \min_{\beta} [Y - W\beta]^T[Y - W\beta]$$

and $\tilde{\sigma}^2 = \hat{\sigma}^2$ hence

$$F = \frac{\min\limits_{\delta^T\beta=\gamma} [Y - W\beta]^T[Y - W\beta] - \min\limits_{\beta} [Y - W\beta]^T[Y - W\beta]}{s\tilde{\sigma}^2}$$

which is the appropriate test in the "β" model. Hence condition (ii) is satisfied and the theorem is proved.

The importance of Theorem 3.0 is that if we are given $Y = W\beta + e$ where W is an $n\times t$ matrix of rank $q < t$, we can transform this to the appropriate "u" model and perform the analysis in the "u" model. Hence we have the ability to work in either model.

Let us now consider another form for the numerator of the test statistic. From Theorem 1.7, we see that

$$\min_{\delta^T\beta=\gamma} [Y - W\beta]^T[Y - W\beta] = [\delta^T\hat{\beta} - \gamma]^T\left[\frac{V[\delta^T\hat{\beta}]}{\sigma^2}\right]^{-1} \cdot [\delta^T\hat{\beta} - \gamma] + R_0^2$$

where

$$R_0^2 = \min_{\beta} [Y - W\beta]^T[Y - W\beta]$$

and

$$V[\delta^T\hat{\beta}] = \sigma^2\delta^T(W^TW)^+\delta$$

Hence

$$\min_{\delta^T\beta=\gamma} [Y - W\beta]^T[Y - W\beta] - \min_{\beta} [Y - W\beta]^T[Y - W\beta]$$

$$= (\delta^T\hat{\beta} - \gamma)^T(\delta^T(W^TW)^+\delta)^{-1}(\delta^T\hat{\beta} - \gamma) = R_1^2 - R_0^2$$

Now we know, by Theorem 3.0, that

$$\lambda^T\hat{u} = \delta^T\hat{\beta}$$

where

$$\delta^T = \lambda^T P \quad .$$

Hence

$$V[\lambda^T \hat{u}] = V[\delta^T \hat{\beta}] \quad .$$

Therefore, we can write

$$R_1^2 - R_0^2 = [\lambda^T \hat{u} - \gamma]^T \left[\frac{V[\lambda^T \hat{u}]^{-1}}{\sigma^2} \right] [\lambda^T \hat{u} - \gamma]$$

but

$$V[\lambda^T \hat{u}] = \sigma^2 \lambda^T A (X^T X)^{-1} \lambda \quad .$$

Hence

$$R_1^2 - R_0^2 = [\lambda^T \hat{u} - \gamma]^T [\lambda^T A (X^T X)^{-1} \lambda]^{-1} [\lambda^T \hat{u} - \gamma] \quad .$$

Thus the test statistic can be written as

$$F = \frac{[\lambda^T \hat{u} - \gamma]^T [\lambda^T A (X^T X)^{-1} \lambda]^{-1} [\lambda^T \hat{u} - \gamma]}{s \hat{\sigma}^2} \quad . \qquad (3.4)$$

Hence we have two forms for the test statistic.

We shall now consider specific models usually encountered in statistics and show how they can be analyzed in the "u" model. One important aspect of using the "u" model should be noted; and that is that there is no need to distinguish between the so called "equal numbers" case and the "unequal numbers" case. The theory developed in Chapter 2 is completely general. If we indeed have equal numbers, then $X^TX = rI$ and all of the formulas simplify. Furthermore, our analysis does not have two sets of equations as do some of the analyses of linear models. For instance, Graybill [15] has a section in which he analyzes the two-way without interaction assuming equal numbers and then a chapter devoted to the analysis when he has unequal numbers.

3.1 Classification Models

In this section, we will consider classification models, that is one-way, two-way without interaction and the two-way with interaction. Extension to higher-way classifications is "easily" seen.

3.1.1 One-Way Classification

Suppose we are given the usual one-way classification, that is

$$y_{ij} = u + \alpha_i + e_{ij} \qquad i = 1,\cdots,a, j = 1,\cdots,n_i$$

Now we can rewrite this in matrix notation as $Y = W\beta + e$, where W is $n\times(a+1)$ of rank a . At this point, we would like to find the appropriate "u" model. From Eq. (3.2), we see that we want θ^T such that $\text{rank}(\theta^T) = \text{rank}(X) - \text{rank}(W)$. But $\text{rank}(X) = a$ and $\text{rank}(W) = a$. Hence $\text{rank}(\theta^T) = 0$ which implies $\theta^T = 0$. Therefore the appropriate "u" model is

$$y_{ij} = u_i + e_{ij} \qquad i = 1,\cdots,a;\ j = 1,\cdots,n_i \quad .$$

In other words the u_i's are unconstrained. Note, however, that we do know that $u_i = u + \alpha_i$ or in matrix notation $u = P\beta$. Therefore let us consider hypotheses about u and α_i in the "β" model. For example, suppose we want to estimate $\alpha_1-\alpha_2$. We need to find a λ^T such that $\lambda^T P = \delta^T$. It is clear that

$$u_1 - u_2 = \lambda^T u = \alpha_1 - \alpha_2 = \delta^T\beta \quad .$$

Hence to estimate $\alpha_1-\alpha_2$ in the "β" model we estimate u_1-u_2 in the "u" model. Since $\theta^T = 0$, $\hat{u}_1 = Y_{1\cdot}/n_1$ and $\hat{u}_2 = Y_{2\cdot}/n_2$. Hence

$$\hat{\alpha}_1 - \hat{\alpha}_2 = \hat{u}_1 - \hat{u}_2 = Y_{1\cdot}/n_1 - Y_{2\cdot}/n_2 \quad ,$$

which is a well-known result. Now suppose we wanted to test H_o: $\alpha_1 = \alpha_2 = \cdots = \alpha_a$. The appropriate test in the "u" model is H_o: $u_1 = u_2 = \cdots = u_a$. We shall present numerical examples of this test and several others for the one-way classification in Chapter 6. The important fact here is to note that given the "β" model $y_{ij} = u + \alpha_i + e_{ij}$, the appropriate "u" model is $y_{ij} = u_i + e_{ij}$.

3.1.2 Two-Way Classification Without Interaction

The usual definition of the two-way classification without interaction model in the "β" model notation is

$$y_{ijk} = u + \alpha_i + \gamma_j + e_{ijk}$$

$i = 1,\cdots,a$; $j = 1,\cdots,b$; $k = 0,1,\cdots,n_{ij}$. In other words, we assume that the model is additive. Graybill [15] gives the following definition of the two-way additive model.

> "The two-way classification model $y_{ij} = u_{ij} + e_{ij}$ will be said to be an additive model if u_{ij} is such that $u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0$ for all i, i', j, j' ."

This definition provides us with the appropriate "u" model; that is $Y = Xu + e$ subject to $\theta^T u = 0$, where a typical row of θ^T is

$$u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0 \ .$$

Now $\text{rank}(\theta^T) = \text{rank}(X) = \text{rank}(XP)$. The rank of X is, say, p and $\text{rank}(XP) = a + b - 1$. Hence, by Theorem 3.0, $\text{rank}(\theta^T) = p - a - b + 1$. Now let us consider various possibilities. First let us suppose $i = 1,\cdots,a$; $j = 1,\cdots,b$, and $k = 1$. Then $p = ab$ and $\text{rank}(\theta^T) = ab - a - b + 1 = (a - 1)(b - 1)$. Also

$$\hat{\sigma}^2 = \frac{1}{n - p + r}[Y - X\hat{u}]^T[Y - X\hat{u}]$$

$$= \frac{1}{ab - ab + r}[Y - \hat{u}]^T[Y - \hat{u}]$$

$$= \frac{1}{(a - 1)(b - 1)}[Y - \hat{u}]^T[Y - \hat{u}]$$

Hence the degrees of freedom for residual are $(a-1)(b-1)$ as is well known for this case.

Second, let us suppose $i = 1,\cdots,a$, $j = 1,\cdots,b$ and $k = 0,1,\cdots,n_{ij}$. In this case, $\text{rank}(X) = p$, which is the number of nonzero n_{ij} . Hence the degrees of freedom for residual are $n_{..}-a-b+1$, where $n_{..} = \sum_i \sum_j n_{ij}$. In this case, θ^T is a $(p-a-b+1)\times p$ matrix.

Let us now consider the structure of θ^T for the above two cases. First, let us consider the following example: Let $i = 1,2,3,4$, $j = 1,2,3$, and $k = 1$. Since a method of constructing θ^T is given in Chapter 5, we will not now discuss the details concerning the construction of θ^T. We will only observe the structure of θ^T. The constraints on the u_{ij} were that $u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0$ for all i, j, i', j'. Hence we have

$$
\begin{aligned}
u_{11} - u_{21} - u_{12} + u_{22} &= 0 \\
u_{11} - u_{21} - u_{13} + u_{23} &= 0 \\
u_{11} - u_{31} - u_{13} + u_{32} &= 0 \\
u_{11} - u_{31} - u_{13} + u_{33} &= 0 \\
u_{11} - u_{41} - u_{12} + u_{42} &= 0 \\
u_{11} - u_{41} - u_{13} + u_{42} &= 0
\end{aligned}
$$

Note that we do not have the constraints of all i, i', j, j'. This is because the above six form a linearly independent set, and any other constraint would be a linear combination of these six and need not be included in θ^T . This is in keeping with the assumption that θ^T be of full-row rank. In matrix notation, we have

$$\theta^T = \begin{array}{c} \begin{array}{cccccccccccc} u_{11} & u_{12} & u_{13} & u_{21} & u_{22} & u_{23} & u_{31} & u_{32} & u_{33} & u_{41} & u_{42} & u_{43} \end{array} \\ \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} \end{array}$$

Let us now consider estimates and tests of hypothesis for this example. The "β" model is

$$y_{ij} = u + \alpha_i + \gamma_j + e_{ij} \quad .$$

The appropriate "u" model is

$$Y = u + e$$

subject to $\theta^T u = 0$ where θ^T is given above. Now we know that

$$u_{ij} = u + \alpha_i + \gamma_j$$

Suppose we wish to estimate $\alpha_i-\alpha_i$, then $u_{ij} - u_{i'j} = \alpha_i - \alpha_{i'}$; or, if we wish to estimate

$\gamma_j - \gamma_{j'}$, we see that $u_{ij} - u_{ij'} = \gamma_j - \gamma_{j'}$. We also see, however, that $\bar{u}_{i\cdot} - \bar{u}_{i'\cdot} = \alpha_i - \alpha_{i'}$ and $\bar{u}_{\cdot j} - \bar{u}_{\cdot j'} = \gamma_j - \gamma_{j'}$. The reason for this is that we have a completely balanced additive model and $u_{i\cdot} - u_{i'\cdot}$ is equivalent to $u_{ij} - u_{i'j}$.

Let us test H_o: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ or

$$H_o: \quad \delta^T\beta = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix} \beta = 0$$

where $\beta^T = [u\ \alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \gamma_1\ \gamma_2\ \gamma_3]$. The appropriate test in the "u" model is H_o: $u_{ij} = u_{i'j}$. In matrix notation, we have:

$$\lambda^T u = \begin{array}{c} \begin{matrix} u_{11} & u_{12} & u_{13} & u_{21} & u_{22} & u_{23} & u_{31} & u_{32} & u_{33} & u_{41} & u_{42} & u_{43} \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \end{array} u$$

From Section 2.3, we see that to test H_o: $\lambda^T u = 0$, we set $B = \begin{bmatrix} \theta^T \\ \lambda^T \end{bmatrix}$, but B is a 15×12 matrix of rank 9. This can be seen by observing that the first row of λ^T , minus the fourth row of λ^T , is equal to the first row of θ^T , and so on. After eliminating the dependent rows of B , we have

$$B = \begin{bmatrix} \theta^T \\ \lambda_1^T \end{bmatrix}$$

where

$$\lambda_1^T = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 \end{bmatrix}$$

or in other words we are testing

$$H_o: \ \bar{u}_{1\cdot} = \bar{u}_{2\cdot} = \bar{u}_{3\cdot} = \bar{u}_{4\cdot} \ .$$

Hence if we wanted to use Eq. (3.4) to calculate F , we would use H_o: $\lambda_1^T u = 0$. Numerical examples of this procedure are given in Chapter 6.

Let us now consider an example where $i = 1,2,3,4$; $j = 1,2,3$; and $k = 0,1,\cdots,n_{ij}$. Suppose we observed the data in the following manner:

i \ j	1	2	3
1	y_{111} y_{112}		y_{131}
2		y_{221}	y_{221}
3	y_{311}	y_{321}	
4	y_{411}		y_{431} y_{432}

A blank cell means no observations were made at the combination of i and j . Again, the "β" model is $y_{ijk} = u + \alpha_i + \gamma_j + e_{ijk}$. The "u" model is now

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{131} \\ y_{221} \\ y_{231} \\ y_{311} \\ y_{321} \\ y_{411} \\ y_{431} \\ y_{432} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{13} \\ u_{22} \\ u_{23} \\ u_{31} \\ u_{32} \\ u_{41} \\ u_{43} \end{bmatrix} + e$$

subject to $\theta^T u = 0$ or subject to

$$u_{11} - u_{13} - u_{41} + u_{43} = 0$$

$$u_{11} - u_{13} - u_{22} + u_{23} - u_{31} + u_{32} = 0 \quad .$$

Thus if the model is to be additive (that is, no interaction) then the u_{ij} must satisfy the above constraints. Again, we will postpone the discussion of the construction of θ^T until Chapter 5.

Let us now consider estimating $\alpha_i - \alpha_{i'}$. We can use $u_{ij} - u_{i'j}$ where u_{ij} and $u_{i'j}$ exist; that is we do not say that $u_{12} - u_{42}$ is an estimate of $\alpha_1 - \alpha_4$ since u_{12} and u_{42} do not exist in the model. Rather, we would use $u_{13} - u_{43}$. Let us observe the following. First $\bar{u}_{1\cdot} - \bar{u}_{2\cdot}$ is not equal to $\alpha_1 - \alpha_2$ as in the balanced case. This can be seen by observing that

$$\bar{u}_{1\cdot} = \frac{u_{11} + u_{13}}{2} = \frac{2u + 2\alpha_1 + \gamma_1 + \gamma_3}{2}$$

$$\bar{u}_{2\cdot} = (u_{22} + u_{23})/2 = (2u + 2\alpha_2 + \gamma_2 + \gamma_3)/2$$

then

$$\bar{u}_{1\cdot} - \bar{u}_{2\cdot} = (\alpha_1 - \alpha_2) + \frac{\gamma_1 - \gamma_2}{2}$$

Also, suppose we test H_o: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$. Again, we test H_o: $u_{ij} = u_{i'j}$ in the "u" model, or

$$\lambda^T u = \begin{array}{c} \begin{array}{cccccccc} u_{11} & u_{13} & u_{22} & u_{23} & u_{31} & u_{32} & u_{41} & u_{43} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \end{array} u = 0$$

Now we augment λ^T to θ^T and obtain B^T , that is

$$B = \begin{array}{c} \begin{array}{cccccccc} u_{11} & u_{13} & u_{22} & u_{23} & u_{31} & u_{32} & u_{41} & u_{43} \end{array} \\ \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \end{array}$$

After eliminating the dependent rows, we obtain

$$B = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \theta^T \\ \lambda_1^T \end{bmatrix}$$

where

$$\lambda_1^T = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Thus, in order to test H_o: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$, we test H_o: $\lambda_1^T u = 0$ or

$$H_o: \begin{cases} u_{11} - u_{31} = 0 \\ u_{22} - u_{32} = 0 \\ u_{13} - u_{43} = 0 \end{cases}$$

Let us restate the procedure for analyzing $y_{ijk} = u + \alpha_i + \gamma_j + e_{ijk}$:

(1) Find θ^T

(2) $\hat{u} = [I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T](X^TX)^{-1}X^TY$

(3) $\lambda^T\hat{u}$ is B.L.U.E. of $\lambda^T u$

(4) H_o: $\lambda^T u = 0$ can be tested by using the results of Chapter 2 or Theorem 3.0.

In concluding this section, let us observe the following. First, we have given estimates and tests of hypothesis for the "classical" estimates and tests of hypothesis; however, there is no need to restrict ourselves just to these. We would have tested, for example, H_o: $2u_{1\cdot} = u_{2\cdot}$ in either of the above cases. As we know, we can test H_o: $\lambda^T u = \xi$ where λ^T is arbitrary and ξ is known.

Second, if we are in the completely balanced case, we know that to test H_o: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ and H_o: $\gamma_1 = \gamma_2 = \gamma_3$, we can use

$$SS(\alpha_i) = \frac{\sum Y_{i\cdot}^2}{3} - \frac{(Y_{\cdot\cdot})^2}{12}$$

$$SS(\gamma_j) = \frac{\sum Y_{\cdot j}^2}{4} - \frac{(Y_{\cdot\cdot})^2}{12}$$

to obtain the numerator sum of squares for the test statistics and we would obtain $\hat{\sigma}^2$ by

$$\hat{\sigma}^2 = \frac{\sum y_{ij}^2 - SS(\alpha_i) - SS(\gamma_i)}{6}$$

In other words, Theorem 3.0 says that $H_o: \lambda^T u = 0$ is equivalent to $H_o: \delta^T \beta = 0$. Thus if the computational form of $SS(\delta^T)$ is simple, we could use it to obtain $SS(\lambda^T)$. If δ^T is not one of the "classical" hypotheses, then simple forms probably do not exist for $SS(\delta^T)$ and we would use $SS(\lambda^T)$ directly.

The generalization to the N-way classification follows similar reasoning and will not be given here.

3.1.3 The Two-Way Classification with Interaction

The usual two-way classification with interaction model is given by $y_{ijk} = u + \alpha_i + \gamma_j + \delta_{ij} + e_{ijk}$ $i = 1,\cdots,a;\ j = 1,\cdots,b;\ k = 0,1,2,\cdots,n_{ij}$. In this case, the rank of X is the number of nonzero n_{ij} , which we denote by p . Also, the rank (XP) is p . Hence by Theorem 3.0,

$$\text{rank}(\theta^T) = \text{rank}(X) - \text{rank}(XP) = p - p = 0 .$$

Hence $\theta^T = 0$. Therefore, the appropriate "u" model is $Y = Xu + e$. Let us first consider the case $n_{ij} = t$; that is, t observations per population. Now the test for interaction in the "β" model is H_o: $\delta_{ij} - \delta_{i'j} - \delta_{ij'} - \delta_{i'j'} = 0$ for all i, i', j, j' . We know, however, that $u_{ij} = u + \alpha_i + \gamma_j + \delta_{ij}$. Hence H_o: $u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0$ for all i, i', j, j' is the appropriate test for no interaction in the "u" model.

Now let us consider $u_{1.}-u_{2.}$. Now

$$\begin{aligned} u_{1.} - u_{2.} &= bu + b\alpha_1 + \gamma_. + \delta_{1.} - bu \\ &\quad - b\alpha_2 - \gamma_. - \delta_{2.} \\ &= b(\alpha_1 - \alpha_2) + \delta_{1.} - \delta_{2.} \end{aligned}$$

or

$$\bar{u}_{1.} - \bar{u}_{2.} = \alpha_1 - \alpha_2 + \frac{\delta_{1.} - \delta_{2.}}{b}$$

If we impose the restrictions $\delta_{i.} = 0$ for all i's, then $\bar{u}_{1.} - \bar{u}_{2.} = \alpha_1 - \alpha_2$. Therefore, in order to test H_o: $\alpha_1 = \alpha_2 = \cdots = \alpha_a$, we test H_o: $\bar{u}_{1.} = \bar{u}_{2.} = \cdots = \bar{u}_{a.}$ in the "u" model. An analogous result is true for the γ_i's. Again, we need not restrict ourselves just to these three tests. Also, for this simple case, we would use the known computational equations to obtain the sums of squares for these three hypotheses.

Let us consider an example where $k = 0,1,\cdots,n_{ij}$. Let $i = 1,2,3,4$; $j = 1,2,3$, and suppose we observe the following:

$i \backslash j$	1	2	3
1	y_{111} y_{112}	y_{121}	y_{131}
2	y_{211}	y_{221} y_{222}	
3	y_{311}	y_{321}	y_{331}
4	y_{411}	y_{421}	y_{431}

In matrix notation, we have

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{131} \\ y_{211} \\ y_{221} \\ y_{222} \\ y_{311} \\ y_{321} \\ y_{331} \\ y_{411} \\ y_{421} \\ y_{431} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{21} \\ u_{22} \\ u_{31} \\ u_{32} \\ u_{33} \\ u_{41} \\ u_{42} \\ u_{43} \end{bmatrix} + e$$

In this case, the test for interaction is H_o: $\lambda^T u = 0$ where

$$\lambda^T = \begin{array}{c} \begin{array}{ccccccccccc} u_{11} & u_{12} & u_{13} & u_{21} & u_{22} & u_{31} & u_{32} & u_{33} & u_{41} & u_{42} & u_{43} \end{array} \\ \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} \end{array}$$

Now let us consider $\bar{u}_{1\cdot} - \bar{u}_{2\cdot}$. We see that

$$\bar{u}_{1\cdot} = u + \alpha_1 + \frac{\gamma_{\cdot}}{3} + \frac{\delta_{11} + \delta_{12} + \delta_{13}}{3}$$

$$\bar{u}_{2\bullet} = u + \alpha_2 + \frac{\gamma_1 + \gamma_2}{2} + \frac{\delta_{21} + \delta_{22}}{2}$$

Now

$$\bar{u}_{1\bullet} - \bar{u}_{2\bullet} = \alpha_1 - \alpha_2 + \frac{\gamma_\bullet}{3} - \frac{(\gamma_1 + \gamma_2)}{2}$$

$$+ \frac{\delta_{1\bullet}}{3} - \frac{(\delta_{21} + \delta_{22})}{2}$$

Suppose $\delta_{1\bullet} = 0$; $\delta_{21} + \delta_{22} = 0$ and $\gamma_\bullet = 0$, then $\bar{u}_{1\bullet} - \bar{u}_{2\bullet} = \alpha_1 - \alpha_2 + \frac{\gamma_3}{2}$. Hence if we test H_o: $\bar{u}_{1\bullet} - \bar{u}_{2\bullet} = 0$, we are testing

$$H_o: \ \alpha_1 - \alpha_2 + \frac{\gamma_3}{2} = 0$$

in the "β" model. Now let us consider

$$\frac{u_{11} + u_{12}}{2} - \frac{u_{21} + u_{22}}{2}$$

Here we see

$$\frac{u_{11} + u_{12}}{2} = u + \alpha_1 + \frac{\gamma_1 + \gamma_2}{2} + \frac{\delta_{11} + \delta_{12}}{2}$$

$$\frac{u_{21} + u_{22}}{2} = u + \alpha_2 + \frac{\gamma_1 + \gamma_2}{2} + \frac{\delta_{21} + \delta_{22}}{2}$$

Hence

$$\frac{u_{11} + u_{12}}{2} - \frac{(u_{21} + u_{22})}{2} = \alpha_1 - \alpha_2 + \frac{\delta_{11} + \delta_{12} - (\delta_{21} + \delta_{22})}{2}$$

Now suppose we set $\delta_{11} + \delta_{12} = 0$ and $\delta_{21} + \delta_{22} = 0$. Then we have that

$$\frac{u_{11} + u_{12}}{2} - \frac{u_{21} + u_{22}}{2} = \alpha_1 - \alpha_2 \quad .$$

Thus, if we test

$$H_o\colon \frac{u_{11} + u_{12}}{2} = \frac{u_{21} + u_{22}}{2} = \frac{u_{31} + u_{32}}{2} = \frac{u_{41} + u_{42}}{2}$$

and impose the proper nonestimable conditions, we will test $H_o\colon \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$.

However, what have we really done? We have ignored all y_{ijk} with $j = 3$. In other words we have "thrown away" data so as to get a complete design and then analyzed it as though $j = 1,2$. If we do this, the interpretation of the results should reflect the fact that $j = 3$ was not included in the analysis. We shall postpone the discussion of the merits of such hypotheses to Chapter 5. The point is that if such a test is desired, it can be done.

Let us summarize the analysis of this model.

(1) $\theta^T = 0$

(2) $\hat{u} = u^* = (X^TX)^{-1}X^TY$

(3) $\lambda^T\hat{u} = \lambda^T(X^TX)^{-1}X^TY$

(4) If H_o: $\lambda^Tu = \xi$, then

$$SS(\lambda^T,\xi) = [\lambda^T\hat{u} - \xi]^T(\lambda^T(X^TX)^{-1}\lambda)^{-1}[\lambda^T\hat{u} - \xi]$$

This concludes the discussion of classification models. Extension to higher-order designs are straightforward.

3.2 Design Models

In this section we will consider some of the more common designs found in classical experimental design theory. The designs in this section include the Latin Square, Balanced Incomplete Block, and the Split Plot Designs. Note that the Randomized Complete Block is not included since the analysis of it is identical to the two-way without interaction. Also, no structure will be assumed about the "treatments" since it would be a simple matter to perform an analysis given a specific structure for the treatments.

3.2.1 Latin Square

The model for the Latin Square is

$$y_{ijk} = u + \alpha_i + \gamma_j + \delta_k + e_{ijk} \quad .$$

We can consider this as a three-way classification without interaction and with missing cells.

We know that there are m^2 observations. Now $X = I$ and $\text{rank}(X) = m^2$. Also $\text{rank}(XP) = 3m - 2$. Hence by Theorem 3.0,

$$\text{rank}(\theta^T) = m^2 - 3m + 2 = (m - 1)(m - 2) \quad .$$

Now from Chapter 2, we know that the degrees of freedom for error are n-p+r, where n is the number of observations, p is the number of populations, and r is the rank (θ^T). Now $n = m^2$, $p = m^2$, $r = (m - 1)(m - 2)$ hence the degrees of freedom for error are

$$n - p + r = (m - 1)(m - 2)$$

which is a well-known result.

Let us consider the following 3×3 Latin Square.

		Column 1	Column 2	Column 3
Row	1	A	B	C
	2	B	C	A
	3	C	A	B

in the "u" notation, we have

u_{111}	u_{122}	u_{133}
u_{212}	u_{223}	u_{231}
u_{313}	u_{321}	u_{332}

The rank of $\theta^T = (m-1)(m-2) = 2 \cdot 1 = 2$. Now

$$\theta^T = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 & -1 & 0 & 1 \\ 1 & 0 & -1 & -1 & 1 & 0 & 0 & -1 & 1 \end{bmatrix} u$$

where

$$u^T = [u_{111}\ u_{122}\ u_{133}\ u_{212}\ u_{223}\ u_{231}\ u_{313}\ u_{321}\ u_{332}]$$

The method of constructing θ^T for the Latin Square is given in Chapter 5. Now the appropriate "u" model is

$$Y = u + e$$

subject to $\theta^T u = 0$.

Let us consider $\bar{u}_{1\cdot\cdot} - \bar{u}_{2\cdot\cdot}$, where $\bar{u}_{1\cdot\cdot}$ means we average over only those terms in the model, that is

$$\bar{u}_{1\cdot\cdot} = \frac{u_{111} + u_{122} + u_{133}}{3}$$

$$u_{2\cdot\cdot} = \frac{u_{212} + u_{223} + u_{231}}{3}$$

Now $u_{ijk} = u + \alpha_i + \gamma_j + \delta_k$. Hence

$$\bar{u}_{1\cdot\cdot} - \bar{u}_{2\cdot\cdot} = \alpha_1 - \alpha_2$$

Thus if we test H_o: $\bar{u}_{1\cdot\cdot} = \bar{u}_{2\cdot\cdot} = \bar{u}_{3\cdot\cdot}$, this is equivalent to H_o: $\alpha_1 = \alpha_2 = \alpha_3$. Likewise, H_o: $\bar{u}_{\cdot 1 \cdot} = \bar{u}_{\cdot 2 \cdot} = \bar{u}_{\cdot 3 \cdot}$ is equivalent to H_o: $\gamma_1 = \gamma_2 = \gamma_3$. And H_o: $\bar{u}_{\cdot\cdot 1} = \bar{u}_{\cdot\cdot 2} = \bar{u}_{\cdot\cdot 3}$ is equivalent to H_o: $\delta_1 = \delta_2 = \delta_3$. Now

$$\bar{u}_{\cdot\cdot 1} = \frac{u_{111} + u_{231} + u_{321}}{3}$$

$$\bar{u}_{\cdot\cdot 2} = \frac{u_{122} + u_{212} + u_{332}}{3}$$

$$\bar{u}_{\cdot\cdot 3} = \frac{u_{313} + u_{223} + u_{133}}{3}$$

Or in matrix notation, the test H_o: $\bar{u}_{\cdot\cdot 1} = \bar{u}_{\cdot\cdot 2} = \bar{u}_{\cdot\cdot 3}$ is:

$$\frac{1}{3}\begin{bmatrix} 1 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & -1 \\ 1 & 0 & -1 & 0 & -1 & 1 & -1 & 1 & 0 \end{bmatrix} u = 0$$

Now let us suppose that u_{111} is missing. In this case, $\text{rank}(\theta^T) = 1$ and

$$\theta^T = [1 \quad -1 \quad -1 \quad 0 \quad 1 \quad 1 \quad -1 \quad 0]u$$

where

$$u^T = [u_{122}\ u_{133}\ u_{212}\ u_{223}\ u_{231}\ u_{313}\ u_{321}\ u_{332}]$$

In this case

$$\bar{u}_{\cdot\cdot 1} = \frac{u_{231} + u_{321}}{2}$$

and

$$\bar{u}_{\cdot\cdot 2} = \frac{u_{122} + u_{212} + u_{332}}{2}$$

So that we see

$$\bar{u}_{\cdot\cdot 1} - \bar{u}_{\cdot\cdot 2} = \delta_1 - \delta_2 + \frac{(\alpha_2 + \alpha_3)}{2} - \frac{\alpha_\cdot}{3} + \frac{(\gamma_2 + \gamma_3)}{2} - \frac{\gamma_\cdot}{3}$$

.

Therefore, H_o: $\bar{u}_{\cdot\cdot 1} - \bar{u}_{\cdot\cdot 2} = 0$ is equivalent to

$$H_o: \ \delta_1 - \delta_2 + \frac{(\alpha_2 + \alpha_3)}{2} - \frac{\alpha_\cdot}{3} + \frac{(\gamma_2 + \gamma_3)}{2} - \frac{\gamma_\cdot}{3} = 0$$

in the "β" model. Now consider

$$H_o: \ \left\{ \begin{array}{l} u_{122} + u_{231} - u_{223} - u_{133} = 0 \\ u_{223} + u_{332} - u_{231} - u_{321} = 0 \end{array} \right\}$$

this is equivalent to

$$H_o: \ \left\{ \begin{array}{l} \delta_1 + \delta_2 - 2\delta_3 = 0 \\ \delta_2 + \delta_3 - 2\delta_1 = 0 \end{array} \right\}$$

or

$$H_o: \ \left\{ \begin{array}{l} \delta_1 - \delta_2 = 0 \\ \delta_2 - \delta_3 = 0 \end{array} \right\}$$

Again we see the flexibility we have in choosing λ^T in order to test H_o: $\lambda^T u = \xi$. If we have the standard Latin Square and the standard analysis is desired, then we would use the known computational forms for finding the appropriate sum of squares. If we have missing data, the "u" approach offers a straightforward method for the analysis. Numerical examples will be given in Chapter 6.

3.2.2 Balanced Incomplete Block (BIB) Design

The model for the BIB design is

$$y_{ijk} = u + \alpha_i + \gamma_k + e_{ijk}$$

$i = 1,\cdots,a$, $j = 1,\cdots,b$. Now $k = 1$ if *ith* treatment is in the *jth* block and $k = 0$ otherwise. This is merely a two-way without interaction with missing cells. It also reflects the condition where the blocks and treatments have been arranged so that $\alpha_i - \alpha_j$ is estimable. In the "u" model, this is equivalent to saying that the populations have been sampled so that the appropriate test for main effects is

$$H_o\colon\ u_{ij} - u_{i'j} = 0$$

The conditions for a BIB design are

(i) Every block contains k treatments

(ii) Every treatment occurs in r blocks

(iii) Every pair of treatments occurs together in the same number (denoted by λ) of blocks.

The number of populations is bk, and there is one observation per population. Hence $X = I$. From Theorem 3.0, we know that

$$\begin{aligned} \mathrm{rank}(\theta^T) &= \mathrm{rank}(X) - \mathrm{rank}(XP) \\ &= \mathrm{rank}(I) - \mathrm{rank}(P) \\ &= bk - a - b + 1 \end{aligned}$$

Let us consider the following BIB design.

Block	Treatment Number
1	1 2
2	1 3
3	2 3

In the "u" model notation, we have

i \ j	1	2	3
1	u_{11}	u_{12}	
2	u_{21}		u_{23}
3		u_{32}	u_{33}

If we had no missing cells, then the constraints on the u_{ij}'s would be $u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0$, for all i, i', j, j' . Let us list the constraints assuming no missing cells. One set of four linearly independent constraints is:

$$u_{11} - u_{21} - u_{12} + u_{22} = 0 \qquad (1)$$

$$u_{11} - u_{21} - u_{13} + u_{23} = 0 \qquad (2)$$

$$u_{11} - u_{31} - u_{12} + u_{32} = 0 \qquad (3)$$

$$u_{11} - u_{31} - u_{13} + u_{33} = 0 \qquad (4)$$

(The procedure for listing this set is given in Chapter 5.) Now u_{13}, u_{22}, u_{31} are missing. Thus we must eliminate these from the above equations.

Subtracting Eq. (2) from Eq. (1) we have

$$-u_{12} + u_{22} + u_{13} - u_{23} = 0$$

and subtracting Eq. (4) from Eq. (3) we have

$$-u_{12} + u_{32} + u_{13} - u_{33} = 0$$

or

$$u_{13} + u_{22} = u_{23} + u_{12}$$

$$u_{13} = u_{33} + u_{12} + u_{32}$$

or $$u_{22} = u_{23} - u_{33} + u_{32}$$

Substituting this into Eq. (1) we have

$$u_{11} - u_{21} - u_{12} + u_{23} - u_{33} + u_{32} = 0$$

Hence

$$\theta^T u = u_{11} - u_{21} - u_{12} + u_{23} - u_{33} + u_{32} = 0 .$$

And $bk - a - b + 1 = 6 - 3 - 3 + 1 = 1$ which is the rank of θ^T. Hence the corresponding "u" model is $Y = u + e$ subject to $\theta^T u = 0$.

Let us now consider a number of hypotheses

(i) H_o: $u_{ij} - u_{i'j} = 0$

(ii) H_o: $u_{i\cdot} - u_{i'\cdot} = 0$

(iii) H_o: $u_{ij} - u_{ij'} = 0$

(iv) H_o: $u_{\cdot j} - u_{\cdot j'} = 0$

As before, we mean by u_{ij}, $u_{i'j}$ and $u_{ij'}$ only those u_{ij}'s that exist in the model, and $u_{i\cdot}$ and $u_{\cdot j}$ mean we sum over those u_{ij} that receive treatment "i" or that are in block j. For example, $u_{1\cdot} = u_{11} + u_{21}$ and $u_{\cdot 1} = u_{12} + u_{32}$, and so on.

The corresponding hypotheses in the "β" model are:

(i) Treatment adjusted for blocks

(ii) Treatments unadjusted

(iii) Blocks adjusted for treatments

(iv) Blocks unadjusted

A numerical example of the BIB design is given in Chapter 6.

3.2.3 Split Plot Design

The model for the Split Plot Design is

$$y_{ijk} = u + \beta_i + \alpha_j + \delta_{ij} + \gamma_k + t_{jk} + e_{ijk}$$

where δ_{ij} and e_{ijk} are assumed to be independent, normally distributed, random variables. If we consider that δ_{ij} is fixed, this is nothing more than a three-way with interaction. This is assuming there is no interaction between α_i and γ_k and that there is no three-way interaction. In this case, the appropriate "u" model is $Y = u + e$ subject to $\theta^T u = 0$, where θ^T is chosen so as to indicate that the model is partially additive. For example, suppose we had the following Split Plot

$$y_{ijk} = u + \beta_i + \alpha_j + \delta_{ij} + \gamma_k + t_{jk} + e_{ijk}$$

where $i = 1,2$, $j = 1,2$, $k = 1,2$ and β_i = the block effect, α_j is the main treatment effect, and γ_k is the subtreatment effect.

Block 1

	Treat. 1	Treat. 2
Sub 1	x	x
Sub 2	x	x

Block 2

	Treat. 1	Treat. 2
Sub 1	x	x
Sub 2	x	x

In this case, we have $\theta^T u = 0$ where

$$\theta^T u =$$

$$\begin{bmatrix} u_{111} - u_{112} - u_{121} + u_{122} - u_{211} + u_{212} + u_{221} - u_{222} \\ u_{111} - u_{112} + u_{121} - u_{122} - u_{211} + u_{212} - u_{221} + u_{222} \end{bmatrix}$$

Again, we refer to Chapter 5 for the construction of θ^T .

The following table gives the appropriate tests in the "u" model

Source of Variation	$H_o: \lambda^T u = 0$
"Blocks"	$u_{i\cdot\cdot} = u_{i'\cdot\cdot}$
"A"	$u_{\cdot j\cdot} = u_{\cdot j'\cdot}$
Error a	$u_{ij\cdot} - u_{i'j\cdot} - u_{ij'\cdot} + u_{i'j'\cdot} = 0$
"B"	$u_{\cdot\cdot k} = u_{\cdot\cdot k'}$
"AB"	$u_{\cdot jk} - u_{\cdot j'k} - u_{\cdot jk'} + u_{\cdot j'k'} = 0$
Error b	$\theta^T u = 0$

Note that

$$\text{error b} = \min_{\theta^T u=0} [Y - Xu]^T[Y - Xu] .$$

But since $X = I$, $\min [Y - Xu]^T[Y - Xu] = SS(\theta^T)$.

Now suppose u_{111} is missing. In this case, $\theta^T u = u_{121} - u_{122} - u_{221} + u_{222} = 0$. Here θ^T is obtained by eliminating u_{111} for the constraint equations given above; however, the hypotheses given above no longer apply since u_{111} is no longer in the model. As in the case of a missing observation in the Latin

Square, there are several hypotheses that seem appropriate. Again we defer discussion of these hypotheses to Chapter 5.

This ends the discussion of relating the classical experimental design models to the "u" model. If we are given any design model of the form $Y = Xu + e$, subject to $u = P\beta$ or $Y = W\beta + e$, we know by Theorem 3.0 how to obtain the appropriate "u" model. The discussion so far has been limited strictly to design models. Through minor modifications, however, we can use the "u" model to describe a regression model or a covariance model. This will also be covered in Chapter 5.

CHAPTER 4. THE MIXED MODEL

4.0 Preliminaries

We now consider the analysis of the "mixed" model — a model in which some of the components are fixed and some are random. Up to now, we have assumed that all the components were fixed, that is we considered u_{ij} as a constant. We will still assume that we are in a design model, that is X is a matrix of zeros and ones. The regression model and covariance model will be considered in Chapter 5.

In order to analyze the mixed model, we will: (1) assume that u_{ij} is a constant; (2) find the sum of squares associated with various hypotheses; (3) assume the proper structure on the u_{ij}, that is we will assume that $u_{ij} \sim N(u^*_{ij}, \sigma^2_{ij})$; and (4) find the expectations of the sum of squares and then test hypothesis or estimate the components of variance. This is the classical method of analyzing mixed models. Here is an example to clarify these concepts. Suppose we have $y_{ijk} = u_{ij} + e_{ijk}$ where

$$u_{ij} = u + \alpha_i + \gamma_j + \delta_{ij}$$

and where

u is a constant

α_i is a constant

$\gamma_j \sim N\left(0, \sigma_\gamma^2\right)$

$\delta_{ij} \sim N\left(0, \sigma_\delta^2\right)$

$e_{ijk} \sim N(0, \sigma^2)$

We would analyze $Y = Xu + e$ assuming u is a constant, that is we would test H_o: $\Lambda^T u = \nu$, etc. Let us denote the sum of squares associated with H_o: $\Lambda^T u = \nu$ by $SS(\Lambda^T, \nu)$. If we are in a balanced case, then the distribution of $SS(\Lambda^T, \nu)$ can be found and a test can be performed. If we are in the unbalanced case, we would estimate σ_γ^2 and σ_δ^2 and find variances of these estimates.

Since the distribution of $SS(\Lambda^T, \nu)$ is known only in the balanced case of the mixed model, we shall not discuss the test of hypotheses in detail. The more important aspects of the mixed model analysis are the problems of estimating variance components and then finding variances of these estimates. We shall consider these problems in depth.

In order to do this, we shall develop a procedure for finding expectations, variances and covariances for sums of squares. This procedure is designed for sum of squares obtained by analyzing the "u" model.

If we test H_o: $\Lambda^T u = \nu$, then the form of the sum of squares, $SS(\Lambda^T,\nu)$, is

$$SS(\Lambda^T,\nu) = [\Lambda^T\hat{u} - \nu]^T[\Lambda^T A(X^TX)^{-1}A^T\Lambda]^{-1} \cdot [\Lambda^T\hat{u} - \nu] \tag{4.1}$$

where A, $(X^TX)^{-1}$ and $\hat{u}$ are defined in Chapter 2.

In order to find a simple expression for Eq. (4.1), we must consider the following theorems.

<u>Theorem 4.0</u> — Suppose we want to test H_o: $\Lambda^T u = \nu$. If T is a nonsingular matrix, then H_o: $T\Lambda^T u = T\nu$ is equivalent to testing H_o: $\Lambda^T u = \nu$, that is $SS(T\Lambda^T,T\nu) = SS(\Lambda^T,\nu)$.

<u>Proof</u> — From Eq. (4.1), we see that

$$\begin{aligned} SS(T\Lambda^T,T\nu) &= [T\Lambda^T\hat{u} - T\nu]^T[T\Lambda^T A(X^TX)^{-1}A^T\Lambda T^T]^{-1} \\ &\quad \cdot [T\Lambda^T\hat{u} - T\nu] \\ &= [\Lambda^T\hat{u} - \nu]^T T^T[T\Lambda^T A(X^TX)^{-1}A^T\Lambda T^T]^{-1}T \\ &\quad \cdot [\Lambda^T\hat{u} - \nu] \end{aligned}$$

$$= [\Lambda^T\hat{u} - \nu]^T[\Lambda^T A(X^TX)^{-1}A^T\Lambda]^{-1}[\Lambda^T\hat{u} - \nu]$$

$$= SS(\Lambda^T,\nu)$$

<u>Theorem 4.1</u> – Suppose we wish to test H_o: $\Lambda^T u = \nu$. There is a Γ such that H_o: $\Gamma^T u = \delta$ is equivalent to H_o: $\Lambda^T u = \nu$, where $\Gamma^T\Gamma = I$ and $\Gamma^T A(X^TX)^{-1}A^T\Gamma = D$, where D is a diagonal matrix.

<u>Proof</u> – By Theorem 1.10, there is a nonsingular matrix B_1 such that $B_1\Lambda^T\Lambda B_1^T = I$. Set $T_1^T = B_1\Lambda^T$. Now by Theorem 1.11, there is an orthogonal matrix B_2 such that

$$B_2T_1^T A(X^TX)^{-1}A^T T_1 B_2^T = D .$$

Let $\Gamma^T = B_2T_1^T = B_2B_1\Lambda^T$. Now B_2B_1 is nonsingular, hence by Theorem 4.0, H_o: $\Gamma^T u = \Gamma^T\nu = \delta$ is equivalent to H_o: $\Lambda^T u = \nu$ where

$$\begin{aligned}\Gamma^T\Gamma &= B_2B_1\Lambda^T\Lambda B_1^T B_2^T \\ &= B_2 I B_2^T \\ &= I\end{aligned}$$

and

$$\Gamma^T A(X^TX)^{-1}A^T\Lambda = B_2T_1^TA(X^TX)^{-1}A^TT_1B_2^T$$

$$= D \quad .$$

The theorem is proved.

From here on if we test H_o: $\Lambda^T u = \nu$, we will assume that Λ^T is such that $\Lambda^T\Lambda = I$ and $\Lambda^T A(X^TX)^{-1}A^T\Lambda = D$. For if Λ^T did not satisfy these conditions, we could find an equivalent hypothesis that does. Hence Eq. (4.1) can be written as

$$\begin{aligned} SS(\Lambda^T,\nu) &= (\Lambda^T\hat{u} - \nu)^T D^{-1}(\Lambda^T\hat{u} - \nu) \\ &= (\Lambda^T\hat{u})^T D^{-1}\Lambda^T\hat{u} - 2\nu^T D^{-1}\Lambda^T\hat{u} \\ &\quad + \nu^T D^{-1}\nu \end{aligned} \tag{4.2}$$

Let us partition Λ^T and ν as follows:

$$\Lambda^T = \begin{bmatrix} \lambda_1^T \\ \hline \vdots \\ \hline \lambda_s^T \end{bmatrix} \qquad \nu = \begin{bmatrix} \nu_1 \\ \hline \vdots \\ \hline \nu_s \end{bmatrix}$$

Hence

$$(\Lambda^T\hat{u})^T D^{-1}(\Lambda^T\hat{u}) = \hat{u}^T \Lambda D^{-1}\Lambda^T\hat{u}$$

$$= \hat{u}^T[\lambda_1 \vdots \cdots \vdots \lambda_s]\begin{bmatrix} \frac{1}{d_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{d_s} \end{bmatrix}$$

$$\cdot \begin{bmatrix} \lambda_1^T \\ \hdashline \vdots \\ \hdashline \lambda_s^T \end{bmatrix}\hat{u}$$

$$= \sum_{i=1}^{s} \frac{1}{d_i}\, \hat{u}^T\lambda_i\lambda_i^T\hat{u} = \sum_{i=1}^{s} \frac{1}{d_i}\left(\lambda_i^T\hat{u}\right)^2$$

Also

$$\nu^T D^{-1}\Lambda^T\hat{u} = \sum_{i=1}^{s} \frac{1}{d_i}\left[\nu_i\left(\lambda_i^T\hat{u}\right)\right]$$

and

$$\nu^T D^{-1}\nu = \sum_{i=1}^{s} \frac{\nu_i^2}{d_i}$$

Hence Eq. (4.2) becomes

$$SS(\Lambda^T,\nu) = \sum_{i=1}^{s} \frac{1}{d_i}\left(\lambda_i^T\hat{u}\right)^2 - 2\sum_{i=1}^{s}\frac{\nu_i\left(\lambda_i^T\hat{u}\right)}{d_i} + \sum_{i=1}^{s}\frac{\nu_i^2}{d_i} \tag{4.3}$$

Now if we test H_o: $\Lambda^T u = 0$, then Eq. (4.3) reduces to

$$SS(\Lambda^T,0) = \sum_{i=1}^{s}\frac{1}{d_i}\left(\lambda_i^T\hat{u}\right)^2 \tag{4.4}$$

We shall denote $SS(\Lambda^T,0)$ by $SS(\Lambda^T)$. Since we normally test hypotheses of the form H_o: $\Lambda^T u = 0$, the remainder of this chapter will give results only for this case. The other case, while not difficult to obtain, involves rather lengthy algebraic expressions.

Let us now find the expected values, variances and covariances of the sum of squares which is given by Eq. (4.4).

4.1 Expectations, Variances and Covariances of Sums of Squares

The following theorem provides us the moments for $SS(\Lambda^T)$.

<u>Theorem 4.2</u> – Let

$$SS\left(\Lambda_1^T\right) = \sum_{i=1}^{s_1} \frac{\left(\lambda_{1i}^T \hat{u}\right)^2}{d_{1i}}$$

and

$$SS\left(\Lambda_2^T\right) = \sum_{j=1}^{s_2} \frac{\left(\lambda_{2j}^T \hat{u}\right)^2}{d_{2j}}$$

where $\hat{u} \sim N(u,R)$, then

$$\text{(i)} \quad E\left[SS\left(\Lambda_1^T\right)\right] = \sum_{i=1}^{s_1} \frac{1}{d_{1i}} E\left[\lambda_{1i}^T \hat{u}\right]^2$$

$$= \sum_{i=1}^{s_1} \frac{1}{d_{1i}} \left[\lambda_{1i}^T R \lambda_{1i} + u^T \lambda_{1i} \lambda_{1i}^T u\right]^2$$

(ii) $$V\left[SS\left(\Lambda_1^T\right)\right] = \sum_{i=1}^{s_1} \frac{1}{d_{1i}^2} V\left[\lambda_{1i}^T \hat{u}\right]^2$$

$$+ 2 \sum_{i}^{s_1} \underset{<}{} \sum_{k} \frac{1}{d_{1i} d_{1k}} \operatorname{Cov}\left[\left(\lambda_{1i}^T \hat{u}\right)^2, \left(\lambda_{1k}^T \hat{u}\right)^2\right]$$

or

$$V\left[SS\left(\Lambda_1^T\right)\right] = \sum_{i=1}^{s_1} \frac{1}{d_{1i}^2} \left[2\left(\lambda_{1i}^T R \lambda_{1i}\right)^2\right.$$

$$\left. + 4u^T \lambda_{1i} \lambda_{1i} R \lambda_{1i} \lambda_{1i}^T u\right]$$

$$+ 2 \sum_{i}^{s_1} \underset{<}{} \sum_{k} \frac{1}{d_{1i} d_{1k}} \operatorname{Cov}\left[\left(\lambda_{1i}^T \hat{u}\right)^2, \left(\lambda_{1k}^T \hat{u}\right)^2\right]$$

(iii) $$\operatorname{Cov}\left[SS\left(\Lambda_1^T\right), SS\left(\Lambda_2^T\right)\right]$$

$$= \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \frac{1}{d_{1i}} \frac{1}{d_{2j}} \operatorname{Cov}\left[\left(\lambda_{1i}^T \hat{u}\right)^2, \left(\lambda_{2j}^T \hat{u}\right)^2\right]$$

(iv) $$\operatorname{Cov}\left[\left(\lambda_{1i}^T \hat{u}\right)^2, \left(\lambda_{2j}^T \hat{u}\right)^2\right] = 2\left(\lambda_{1i}^T R \lambda_{2j}\right)^2$$

$$+ 4u^T \lambda_{2j} \lambda_{2j}^T R \lambda_{1i} \lambda_{1i}^T u$$

Proof — The proof of this theorem relies on repeated application of Theorem 1.8. Now we have that

$$E\left[SS\left(\Lambda_1^T\right)\right] = \sum_{i=1}^{s_1} \frac{1}{d_{1i}} E\left[\lambda_{1i}^T \hat{u}\right]^2$$

From part (i) of Theorem 1.8, we see that

$$\begin{aligned}
E\left[\lambda_{1i}^T \hat{u}\right]^2 &= E\left[\hat{u}^T \lambda_{1i} \lambda_{1i}^T \hat{u}\right] \\
&= \operatorname{Trace}\left[R\lambda_{1i}\lambda_{1i}^T\right] + u^T\lambda_{1i}\lambda_{1i}^T u \\
&= \operatorname{Trace}\left[\lambda_{1i}^T R\lambda_{1i}\right] + u^T\lambda_{1i}\lambda_{1i}^T u \\
&= \lambda_{1i}^T R\lambda_{1i} + u^T\lambda_{1i}\lambda_{1i}^T u
\end{aligned}$$

Hence part (i) is established.

Now

$$V\left[SS\left(\Lambda_1^T\right)\right] = \sum_{i=1}^{s_1} \frac{1}{d_{1i}^2} V\left[\lambda_{1i}^T \hat{u}\right]^2 + C$$

where

$$C = 2 \sum_{i} \underset{<}{} \sum_{k}^{s_1} \frac{1}{d_{1i} d_{1k}} \operatorname{Cov}\left[\left(\lambda_{1i}^T \hat{u}\right)^2, \left(\lambda_{1k}^T \hat{u}\right)^2\right]$$

Again by Theorem 1.8, part (ii), we have that

$$V\left(\lambda_{1i}^T\hat{u}\right)^2 = 2\ \mathrm{Trace}\left[R\lambda_{1i}\lambda_{1i}^T\right]^2 + 4u^T\lambda_{1i}\lambda_{1i}^T R\lambda_{1i}\lambda_{1i}^T u$$

$$= 2\ \mathrm{Trace}\left[R\lambda_{1i}\lambda_{1i}^T R\lambda_{1i}\lambda_{1i}^T\right]$$

$$+ 4u^T\lambda_{1i}\lambda_{1i}^T R\lambda_{1i}\lambda_{1i}^T u$$

$$= 2\ \mathrm{Trace}\left[\lambda_{1i}^T R\lambda_{1i}\lambda_{1i}^T R\lambda_{1i}\right]$$

$$+ 4u^T\lambda_{1i}\lambda_{1i}^T R\lambda_{1i}\lambda_{1i}^T u$$

$$= 2\left(\lambda_{1i}^T R\lambda_{1i}\right)^2 + 4u^T\lambda_{1i}\lambda_{1i}^T R\lambda_{1i}\lambda_{1i}^T u$$

The expression for C will be given later. Hence part (ii) is established. Now part (iii) is established by the linear property of covariance. By Theorem 1.8, part (iii), we see that

$$\mathrm{Cov}\ \left[\left(\lambda_{1i}^T\hat{u}\right)^2, \left(\lambda_{2j}^T\hat{u}\right)^2\right] = 2\ \mathrm{Trace}\left[R\lambda_{1i}\lambda_{1i}^T R\lambda_{2j}\lambda_{2j}^T\right]$$

$$+ 4u^T\lambda_{1i}\lambda_{1i}^T R\lambda_{2j}\lambda_{2j}^T u$$

$$= 2\ \mathrm{Trace}\left[\lambda_{2j}^T R\lambda_{1i}\lambda_{2j}^T R\lambda_{1i}\right]$$

$$+ 4u^T\lambda_{1i}\lambda_{1i}^T R\lambda_{2j}\lambda_{2j}^T u$$

$$\mathrm{Cov}\left[\left(\lambda_{1i}^T\hat{u}\right)^2,\left(\lambda_{2j}^T\hat{u}\right)^2\right] = 2\left(\lambda_{2j}^T R\lambda_{1i}\right)^2 + 4u^T\lambda_{1i}\lambda_{1i}^T R\lambda_{2j}\lambda_{2j}^T u$$

Hence part (iv) is established. Now C in part (ii) is seen to be

$$C = 2\sum_{i}^{s_1}\sum_{<\ k}\frac{1}{d_{1i}d_{1k}}\left[2\left(\lambda_{1k}^T R\lambda_{1i}\right)^2 + 4u^T\lambda_{1i}\lambda_{1i}^T R\lambda_{1k}\lambda_{1k}^T u\right]$$

Hence the theorem is proved.

Let us discuss some of the computation aspects of Theorem 4.2. Let $b_{1i} = R\lambda_{1i}$, $b_{2j} = R\lambda_{2j}$, $f_{1i} = u^T\lambda_{1i}$ and $f_{2j} = u^T\lambda_{2j}$. These quantities are needed for $E\left[SS\left(\Lambda_1^T\right)\right]$ and $E\left[SS\left(\Lambda_2^T\right)\right]$, that is

$$E\left[SS\left(\Lambda_1^T\right)\right] = \sum_{i=1}^{s_1}\frac{1}{d_{1i}}\left[\lambda_{1i}^T b_{1i} + f_{1i}^2\right]^2$$

But now we see that these quantities are needed for $V\left[SS\left(\Lambda_1^T\right)\right]$, that is

$$V\left[SS\left(\Lambda_1^T\right)\right] = \sum_{i=1}^{s_1}\frac{1}{d_{1i}^2}\left[2\left(\lambda_{1i}^T b_{1i}\right)^2 + 4f_{1i}^2\lambda_{1i}^T b_{1i}\right]$$

$$+ 2\sum_{i}^{s_1}\sum_{<\ k}\frac{1}{d_{1i}d_{1k}}\left[2\left(\lambda_{1i}^T b_{1k}\right)^2 + 4f_{1j}f_{1k}\lambda_{1i}^T b_{1k}\right]$$

and

$$\operatorname{Cov}\left[SS\left(\Lambda_1^T\right), SS\left(\Lambda_2^T\right)\right] = \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \frac{1}{d_{1i} d_{2j}} \left[2\left(\lambda_{1i}^T b_{2j}\right)^2 + 4 f_{1i} f_{2j} \lambda_{1i}^T b_{2j}\right]$$

Once the expectations are found, the effort to find the variances and covariances is minimal.

Before returning to the analysis of the mixed model, let us note that in the completely fixed model, there is no need to find $E\left[SS\left(\Lambda_1^T\right)\right]$ and $V\left[SS\left(\Lambda_1^T\right)\right]$ by Theorem 4.2 because $SS\left(\Lambda_1^T\right)/s_1$ is a noncentral chi-square and the moments can be expressed in terms of the noncentrality parameter.

From Eq. (2.9), we see that

$$\hat{u} = u + A(X^T X)^{-1} X^T e$$

Now we will assume a structure on the u_{ij}, that is we will write $Y = Xu + e$ subject to $\theta^T u = 0$ where

$$u = \sum_{i=1}^{k_1} U_i \alpha_i + \sum_{j=1}^{k_2} V_j b_j \tag{4.5}$$

and where

U_i is a $p \times s_i$ matrix associated with α_i

α_i is an $s_i \times 1$ constant vector

V_j is a $p \times s_j$ matrix associated with b_j

b_j is an $s_j \times 1$ random vector such that $b_j \sim N\left(0, \sigma^2_{b_j} I\right)$

$e \sim N(0, \sigma^2 I)$

e, b_j and $b_{j'}$ for $j \neq j'$ are mutually independent.

Now we see that

$$\hat{u} = \sum_{i=1}^{k_1} U_i \alpha_i + \sum_{j=1}^{k_2} V_j b_j + A(X^T X)^{-1} X^T e$$

Hence

$$E[\hat{u}] = \sum_{i=1}^{k_1} U_i \alpha_i$$

$$V[\hat{u}] = \sum_{j=1}^{k_2} V_j V_j^T \sigma^2_{b_j} + A(X^T X)^{-1} A^T \sigma^2$$

Let us now show that $SS(\Lambda^T)$ and $\hat{\sigma}^2$ are independent where

$$\hat{\sigma}^2 = \frac{(Y - X\hat{u})^T(Y - X\hat{u})}{n - p + r}$$

$$= \frac{e^T We}{n - p + r}$$

where $W = I - XA(X^TX)^{-1}X^T$. Now we know that $\hat{u} - u = A(X^TX)^{-1}X^Te$. Let $z = We$. Hence $\hat{u}-u$ and z are independent since $W^TA(X^TX)^{-1}X^T = 0$. Thus $\mathrm{Cov}\,[\hat{u} - u, z] = 0$. But

$$\mathrm{Cov}\,[\hat{u} - u, z] = \mathrm{Cov}\,[\hat{u},z] - \mathrm{Cov}\,[u,z]$$

But $\mathrm{Cov}\,[u,z] = \mathrm{Cov}\,[u,We] = \mathrm{Cov}\,[u,e]W = 0$ since we are assuming by the structure of u that u and e are independent. Hence $\mathrm{Cov}\,[\hat{u},z] = 0$. But $\hat{u}$ and z are both normal, hence $\hat{u}$ and z are independent. Hence $SS\left(\Lambda_1^T\right)$ and $\hat{\sigma}^2$ are independent. This fact will be useful when we are finding variances of the estimates of components of variance. This results says that the sum of squares associated with H_o: $\Lambda^Tu = 0$ is independent of $\hat{\sigma}^2$ no matter if the model is fixed or random, or if we have a balanced or unbalanced design.

At this point, there are sufficient tools to find estimates of components of variance and the variance of these estimates. This can be done by using Theorem 4.2 and assuming the proper structure on u, that is u has a structure as given by Eq. (4.5).

In order to provide a better understanding of these concepts, we shall consider a number of specific examples.

4.2 Mixed Model Examples

4.2.1 The One-Way Classification

For this model, we assume

$$y_{ij} = u_i + e_{ij} \qquad i = 1,\cdots,a,a+1; \quad j = 1,\cdots,n_i$$

where $u_i = \gamma + b_i$ and $b_i \sim N\left(0,\sigma_b^2\right)$. In terms of Eq. (4.5), we have that $u = \gamma 1 + V_1 b$. Now suppose we wish to estimate σ_b^2. We begin by assuming a completely fixed model. Let us test H_o: $u_i = u_{i'}$. Then $SS(\Lambda^T) = \sum \frac{1}{d_i}\left(\lambda_i^T\hat{u}\right)^2$. But now from Eq. (2.9), we know that

$$\hat{u} = \gamma 1 + V_1 b + A(X^TX)^{-1}X^Te$$

Now for the one-way classification, we know that $V_1 = I$ and $A = I$, hence

$$\hat{u} = \gamma 1 + b + (X^T X)^{-1} X^T e \qquad (4.6)$$

Thus $E[\hat{u}] = \gamma 1$, and

$$V[\hat{u}] = I \cdot \sigma_b^2 + \sigma^2 (X^T X)^{-1}$$

Therefore, we see by Theorem 4.2 that

$$E[SS(\Lambda^T)] = \sum_{i=1}^{a} \frac{1}{d_i} \left(\lambda_i^T R \lambda_i + \gamma^2 (1^T \lambda_i)^2 \right)$$

But $1^T \lambda_i = 0$ since we are testing

$$H_o: \; u_1 = u_2 = \cdots = u_a = u_{a+1}$$

Also we see that

$$\lambda_i^T R \lambda_i = \sigma_b^2 \lambda_i^T \lambda_i + \lambda_i^T (X^T X)^{-1} \lambda_i \sigma^2$$

$$= \sigma_b^2 + d_i \sigma^2$$

since $\Lambda^T (X^T X)^{-1} \Lambda = D$. Therefore

$$E[SS(\Lambda^T)] = a\sigma^2 + \sigma_b^2 \sum_{i=1}^{a} \frac{1}{d_i}$$

Now let

$$k_0 = \sum_{i=1}^{a} \frac{1}{d_i}$$

Let us use $\hat{\sigma}^2$ from Theorem 2.2 as an estimate of σ^2. Hence an unbiased estimate of σ_b^2 is

$$\hat{\sigma}_b^2 = \frac{SS(\Lambda^T) - a\hat{\sigma}^2}{k_0} \tag{4.7}$$

Now

$$V\left[\hat{\sigma}_b^2\right] = \left[V[SS(\Lambda^T)] + a^2 V(\hat{\sigma}^2) - 2a \text{ Cov } [SS(\Lambda^T), \hat{\sigma}^2]\right] \Big/ k_0^2$$

and

$$V[SS(\Lambda^T)] = \sum_{i=1}^{a} \frac{1}{d_i^2} \left[2\left(\lambda_i^T R \lambda_i\right)^2\right] + 2 \sum_{i} \sum_{<\, j}^{a} \left(\lambda_i^T R \lambda_j\right)^2$$

where

$$\lambda_i^T R \lambda_i = \sigma_b^2 + d_i \sigma^2$$

and

$$\lambda_i^T R \lambda_j = 0$$

Thus

$$V[SS(\Lambda^T)] = 2 \sum_{i=1}^{a} \frac{1}{d_i^2} \left[\sigma_b^2 + d_i \sigma^2\right]^2$$

$$= 2\left[\sigma_b^4 \sum_{i=1}^{a} \frac{1}{d_i^2} + 2\sigma_b^2\sigma^2 \sum_{i=1}^{a} \frac{1}{d_i} + a\sigma^4\right]$$

$$= 2\left[k_1\sigma_b^4 + 2k_0\sigma_b^2\sigma^2 + a\sigma^4\right]$$

where

$$k_1 = \sum_{i=1}^{a} \frac{1}{d_i^2}$$

$$k_0 = \sum_{i=1}^{a} \frac{1}{d_i}$$

Now

$$\frac{(n - (a + 1))\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n - (a + 1))$$

hence

$$V(\hat{\sigma}^2) = \frac{2\sigma^4}{(n - (a + 1))}$$

Since $SS(\Lambda^T)$ and $\hat{\sigma}^2$ are independent, we know that $\mathrm{Cov}\,[SS(\Lambda^T),\hat{\sigma}^2] = 0$. Thus

$$V\left[\hat{\sigma}_b^2\right] = \left[2\left(k_1\sigma^4 + 2k_0\sigma_b^2\sigma^2 + a\sigma^4\right) + \frac{2a^2\sigma^4}{(n - (a + 1))}\right] \Bigg/ k_0^2 \tag{4.8}$$

Let us now consider the special case where we have an equal number of observations from each population, that is t observations per cell. In this case, $X^TX = tI$ and $D = (X^TX)^{-1} = \frac{1}{t}\, I$. Hence we see that

$$k_0 = \sum_{i=1}^{a} \frac{1}{d_i} = at$$

$$k_1 = \sum_{i=1}^{a} \frac{1}{d_i^2} = at^2$$

Thus

$$\hat{\sigma}_b^2 = \frac{SS(\Lambda^T) - a\hat{\sigma}^2}{at}$$

and

$$V\left[\hat{\sigma}_b^2\right] = 2\left[at^2\sigma_b^4 + 2at\sigma_b^2\sigma^2 + a\sigma^4 + \frac{a^2}{(n - (a + 1))}\,\sigma^4\right] \Bigg/ a^2t^2$$

$$= \frac{2\sigma_b^4}{a} + \frac{4\sigma_b^2\sigma^2}{at} + \frac{2(n-1)}{(n-(a+1))t^2a}\sigma^4$$

which is the classical result for the variance of $\hat{\sigma}_b^2$ in the balanced case.

Now it is easy to see that Eq. (4.7) represents an unbiased estimate for σ_b^2 for any Λ^T such that $\Lambda^T 1 = 0$. Thus it is possible to generate an entire class of unbiased estimates of σ_b^2 . And Eq. (4.8) is the general expression for the variance of such an estimate. Now if $\Lambda^T 1 \neq 0$, then it would be necessary to obtain an estimate of γ since $E[SS(\Lambda^T)]$ would contain a γ . This does not pose any problem since Theorem 4.2 is completely general.

4.2.2 The Two-Way Classification Model Without Interaction (Random)

For the moment, we will restrict ourselves to finding estimates of the components of variance. We will assume the model is $Y = Xu + e$ subject to $\theta^T u = 0$, where $\theta^T u = 0$ implies an additive model. We will also assume that

$$u = \gamma 1 + V_1 a + V_2 b$$

that is $a \sim N\left(0,\sigma_a^2\right)$ and $b \sim N\left(0,\sigma_b^2\right)$ and a, b, e are independent. Rather than discuss a completely general model, let us take a specific case and obtain results for it. Let us assume

$$y_{ijk} = \gamma + a_i + b_j + e_{ijk}$$

where $i = 1,2,3$, $j = 1,2$ and $k = 1,\cdots,n_{ij}$, where $n_{11} = 2$, $n_{12} = 2$, $n_{21} = 5$, $n_{22} = 3$, $n_{31} = 1$, $n_{32} = 5$. Hence

$$\theta^T u = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} u = 0$$

$$V_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad V_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Suppose we wish to test (1) H_o: $u_{1\cdot} = u_{2\cdot} = u_{3\cdot}$ and (2) H_o: $u_{\cdot 1} = u_{\cdot 2}$. Let Λ_1 denote the appropriate hypothesis for (1), and Λ_2 denote the appropriate hypothesis for (2).

Now

$$A(X^TX)^{-1}A^T =$$

$$\begin{bmatrix} .3174 & .1826 & .0506 & -.0843 & .1124 & -.0225 \\ .1826 & .3174 & -.0506 & .0843 & -.1124 & .0225 \\ .0506 & -.0506 & .1629 & .0618 & .0843 & -.0169 \\ -.0843 & .0843 & .0618 & .2303 & -.1404 & .0281 \\ .1124 & -.1124 & .0843 & -.1404 & .3539 & .1292 \\ -.0225 & .0225 & -.0169 & .0281 & .1292 & .1742 \end{bmatrix}$$

and

$$\Lambda_1^T = \begin{bmatrix} .2074 & .2074 & -.5703 & -.5703 & .3629 & .3629 \\ .5388 & .5388 & -.0898 & -.0898 & -.4490 & -.4490 \end{bmatrix}$$

$$\Lambda_2^T = .4083[1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1]$$

and

$$\Lambda_1^T A(X^TX)^{-1}A^T\Lambda_1 = \begin{bmatrix} .333 & 0 \\ 0 & .449 \end{bmatrix}$$

$$\Lambda_2^T A(X^TX)^{-1}A^T\Lambda_2 = .4045$$

Thus

$$\hat{u} = \gamma 1 + V_1 a + V_2 b + A(X^TX)^{-1}X^Te$$

and

$$E[\hat{u}] = \gamma 1$$

$$V[\hat{u}] = V_1 V_1^T \sigma_a^2 + V_2 V_2^T \sigma_b^2 + A(X^T X)^{-1} A^T \sigma^2$$

and

$$\lambda_{11}^T [V[\hat{u}]] \lambda_{11} = \lambda_{11}^T V_1 V_1^T \lambda_{11} \sigma_a^2 + d_{11} \sigma^2$$

$$\lambda_{12}^T [V[\hat{u}]] \lambda_{12} = \lambda_{12}^T V_1 V_1^T \lambda_{12} \sigma_a^2 + d_{12} \sigma^2$$

$$\lambda_{21}^T [V[u]] \lambda_{21} = \lambda_{21}^T V_2 V_2^T \lambda_{21} \sigma_b^2 + d_{21} \sigma^2$$

Let

$$b_{11}^T = \lambda_{11}^T V_1 = 2[.2074 \quad -.5703 \quad .3629]$$

$$b_{12}^T = \lambda_{12}^T V_1 = 2[.5388 \quad -.0989 \quad -.4490]$$

$$b_{21}^T = \lambda_{21}^T V_2 = 3[.4083 \quad -.4083]$$

and

$$f_{21}^T = \lambda_{21}^T A(X^T X)^{-1} A^T = [.4046 \quad -.4046 \quad .3035$$

$$-.5056 \quad .6742 \quad -.1349]$$

Thus from Theorem 4.2, we see that

$$E\left[SS\left(\Lambda_1^T\right)\right] = \frac{1}{d_{11}}\left[b_{11}^T b_{11}\sigma_a^2 + d_{11}\sigma^2\right]$$

$$+ \frac{1}{d_{12}}\left[b_{12}^T b_{12}\sigma_a^2 + d_{12}\sigma^2\right]$$

$$= \sigma_a^2\left[\frac{b_{11}^T b_{11}}{d_{11}} + \frac{b_{12}^T b_{12}}{d_{12}}\right] + 2\sigma^2$$

$$= 2\sigma^2 + \left(\frac{2}{1/3} + \frac{2}{.449}\right)\sigma_a^2$$

$$= 2\sigma^2 + (6 + 4.45)\sigma_a^2 = 2\sigma^2 + 10.45\sigma_a^2$$

and

$$E\left[SS\left(\Lambda_2^T\right)\right] = \frac{b_{21}^T b_{21}^T}{d_{21}}\sigma_b^2 + \sigma^2 = \frac{3}{.4045}\sigma_b^2 + \sigma^2$$

$$= \sigma^2 + 7.417\sigma_b^2$$

and

$$V\left[SS\left(\Lambda_1^T\right)\right] = \frac{2}{d_{11}^2}\left[\left(b_{11}^T b_{11}\sigma_a^2 + d_{11}\sigma^2\right)^2\right]$$

$$+ \frac{2}{d_{12}^2}\left[\left(b_{12}^T b_{12}\sigma_a^2 + d_{12}\sigma^2\right)^2\right]$$

$$+ \frac{1}{d_{11}}\frac{1}{d_{12}}\left[2\left(b_{11}^T b_{12}\sigma_a^2\right)^2\right]$$

$$V\left[SS\left(\Lambda_1^T\right)\right] = 4\sigma^4 + 2\left[\left(\frac{b_{11}^T b_{11}}{d_{11}}\right)^2 + \left(\frac{b_{12}^T b_{12}}{d_{12}}\right)^2 + \left(\frac{b_{11}^T b_{12}}{d_{11}d_{12}}\right)\right]\sigma_a^4$$

$$+ 4\left[\frac{b_{11}^T b_{11}}{d_{11}} + \frac{b_{12}^T b_{12}}{d_{12}}\right]\sigma_a^2\sigma^2$$

$$= 4\sigma^4 + 2\left(6^2 + (4.45)^2 + \frac{.000018}{.449}\right)\sigma_a^2$$

$$+ 4(10.45)\sigma_a^2\sigma^2$$

$$= 4\sigma^4 + 111.6\sigma_a^2 + 41.80\sigma_a^2\sigma^2$$

$$V[SS(\Lambda_2)] = 2\sigma^4 + 4\left(\frac{b_{21}^T b_{21}}{d_{21}}\right)\sigma_b^2\sigma^2 + 2\left(\frac{b_{21}^T b_{21}}{d_{21}}\right)^2\sigma_b^4$$

$$= 2\sigma^4 + 4(7.417)\sigma_b^2\sigma^2 + 2(7.417)^2\sigma^4$$

$$= 2\sigma^4 + 29.66\sigma_b^2\sigma^2 + 110.02\sigma_b^4$$

and

$$\text{Cov}\left[SS\left(\Lambda_1^T\right), SS\left(\Lambda_2^T\right)\right] = \frac{1}{d_{11}d_{21}}\, 2\left[\left(f_{21}^T\lambda_{11}\sigma^2\right)^2\right] + \frac{1}{d_{12}}\,\frac{2}{d_{21}}\left(f_{21}^T\lambda_{12}\sigma^2\right)^2$$

$$= 2\sigma^4\left[\frac{\left(f_{21}^T\lambda_{11}\right)^2}{d_{11}d_{21}} + \frac{\left(f_{21}^T\lambda_{12}\right)^2}{d_{12}d_{21}}\right]$$

$$= 2\sigma^4(.9938) = 1.988\sigma^4$$

Now

$$\hat{\sigma}_a^2 = \begin{bmatrix} \frac{1}{10.45} & \frac{-2}{10.45} \end{bmatrix} \begin{bmatrix} SS\left(\Lambda_1^T\right) \\ \hat{\sigma}^2 \end{bmatrix}$$

$$\hat{\sigma}_b^2 = \begin{bmatrix} \frac{1}{7.417} & \frac{-1}{7.417} \end{bmatrix} \begin{bmatrix} SS\left(\Lambda_1^T\right) \\ \hat{\sigma}^2 \end{bmatrix}$$

and since $V(\hat{\sigma}^2) = \sigma^4/7$, the variances of $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ can be easily found. Let us now outline the procedure for finding estimates of σ_a^2 and σ_b^2 for arbitrary hypothesis. Suppose we have Λ_1^T and Λ_2^T. Then

$$E\left[SS\left(\Lambda_1^T\right)\right] = k_1\sigma^2 + k_2\sigma_a^2 + k_3\sigma_b^2$$

$$E\left[SS\left(\Lambda_2^T\right)\right] = k_4\sigma^2 + k_5\sigma_a^2 + k_5\sigma_b^2$$

Now if we set

$$\begin{bmatrix} k_1 & k_2 & k_3 \\ k_4 & k_5 & k_6 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma^2 \\ \sigma_a^2 \\ \sigma_b^2 \end{bmatrix} = \begin{bmatrix} SS\left(\Lambda_1^T\right) \\ SS\left(\Lambda_2^T\right) \\ \text{Residual} \end{bmatrix}$$

Then

$$\begin{bmatrix} \hat{\sigma}^2 \\ \hat{\sigma}_a^2 \\ \hat{\sigma}_b^2 \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_4 & k_5 & k_6 \\ 1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} SS\left(\Lambda_1^T\right) \\ SS\left(\Lambda_2^T\right) \\ \text{Residual} \end{bmatrix}$$

and the covariance matrix of the estimates can be found from the covariance matrix of $SS\left(\Lambda_1^T\right)$, $SS\left(\Lambda_2^T\right)$ and residual. The values of k_i and the covariance matrix of the sums of squares can be obtained by Theorem 4.2. Let us observe that σ^2 does not have to be estimated by residual. Such an estimate, however, has many desirable properties, i.e., central chi-square, unbiased and so on.

4.2.3 The Two-Way Classification Without Interaction (Mixed)

Let us now assume that

$$u = \gamma 1 + U_1\alpha + V_1 b \tag{4.9}$$

where γ, α, U_1, V_1, b are as defined above. Before we analyze this model, let us consider the idea of combining unbiased estimates. Suppose $t_1, \cdots, t_n$ are unbiased estimates of θ and that V is the covariance matrix of the t_i's. We would like to find a set of a_i's such that $\sum a_i t_i$ is an unbiased estimate of θ and $V\left(\sum a_i t_i\right)$ is a minimum. Let $t^T = (t_1, \cdots, t_n)$ and $a^T = (a_1, \cdots, a_n)$, then the problem is to minimize $a^T V a$ subject to $E[a^T t] = \theta$ or subject to $a^T \ell = 1$ where ℓ is a vector of all ones. The choice of a, as given by Theorem 1.13, is

$$a = \frac{V^{-1}\ell}{\ell^T V^{-1} \ell}$$

Now suppose V is unknown. If we can estimate V, say by $\hat{V}$, so that $\hat{V}$ and t are uncorrelated, then $a^T t = \dfrac{\ell^T \hat{V}^{-1} t}{\ell^T \hat{V}^{-1} \ell}$ is an unbiased estimate of θ since

$$\begin{aligned} E[a^T t] &= E\left[\frac{\ell^T \hat{V}^{-1}}{\ell^T \hat{V}^{-1} \ell}\right] E[t] = E\left[\frac{\ell^T \hat{V}^{-1}}{\ell^T \hat{V}^{-1} \ell}\right] \ell \theta \\ &= E\left[\frac{\ell^T \hat{V}^{-1} \ell}{\ell^T \hat{V}^{-1} \ell}\right] \theta = \theta \end{aligned}$$

The merits of such an estimate will not be discussed here.

Let us now analyze the u structure as given by Eq. (2.9). In order to obtain estimates of σ_b^2 and σ^2 , we can use the procedure outlined above. The interest now is in estimates of the constants. For example, suppose we wanted to find an unbiased estimate of $\alpha_1 - \alpha_2$ in the "β" model. There are two methods by which we can do this. First, find λ_1^T so that $\lambda_1^T u = \gamma\lambda_1^T 1 + \lambda_1^T U_1 \alpha + \lambda_1^T V_1 b = \alpha_1 - \alpha_2$; and then use $\lambda_1^T \hat{u}$ as an estimate. Second, find a λ_2^T so that $E\left[\lambda_2^T \hat{u}\right] = \alpha_1 - \alpha_2$ and use $\lambda_2^T \hat{u}$ as an estimate. The difference between the two methods is the following: In the first case, $\lambda_1^T \hat{u} = \alpha_1 - \alpha_2 + \lambda_1^T A(X^T X)^{-1} X^T e$ and $E\left[\lambda_1^T \hat{u}\right] = \alpha_1 - \alpha_2$; while in the second case $\lambda_2^T \hat{u} = \alpha_1 - \alpha_2 + \lambda_2^T V_1 b + \lambda_2^T A(X^T X)^{-1} X^T e$, and

$$E\left[\lambda_2^T \hat{u}\right] = \alpha_1 - \alpha_2 + \lambda_2^T V(E(b)) = \alpha_1 - \alpha_2$$

since $E(b) = 0$. Thus in the second case, $\lambda_2^T u \neq \alpha_1 - \alpha_2$ but $E\left[\lambda_2^T \hat{u}\right] = \alpha_1 - \alpha_2$. Hence $\lambda_1^T \hat{u}$ and $\lambda_2^T \hat{u}$ are two estimates of $\alpha_1 - \alpha_2$. Let us consider the following set.

Let $S = \{\lambda / \lambda^T U_1 \alpha = \alpha_1 - \alpha_2 \text{ and } \gamma\lambda^T 1 = 0\}$ then if $\lambda_i \varepsilon S$,

$$E\left[\lambda_i^T \hat{u}\right] = E\left[(\alpha_1 - \alpha_2) + \lambda_i^T V_1 b + \lambda_i^T A(X^T X)^{-1} X^T e\right]$$

$$E\left[\lambda_i^T \hat{u}\right] = \alpha_1 - \alpha_2$$

Now

$$V\left(\lambda_i^T \hat{u}\right) = \left(\lambda_i^T V_1 V_1 \lambda_i\right)\sigma_b^2 + d_i \sigma^2$$

If we can estimate σ_b^2 and σ^2 so that $\lambda_i^T \hat{u}$ is uncorrelated with $\hat{\sigma}_b^2$ or $\hat{\sigma}^2$, then we can obtain an estimate as described above. For illustration let us consider the same example as described in Section 4.2.2. Now one choice for λ_1^T would be

$$\lambda_1^T = [1 \quad 0 \quad -1 \quad 0 \quad 0 \quad 0]$$

and λ_2^T could be

$$\lambda_2^T = [1 \quad 0 \quad 0 \quad -1 \quad 0 \quad 0]$$

Now

$$\lambda_1^T \hat{u} = \alpha_1 - \alpha_2 + \lambda_1^T A (X^T X)^{-1} X^T e$$

and

$$\lambda_2^T \hat{u} = \alpha_1 - \alpha_2 + b_1 - b_2 + \lambda_2^T A (X^T X)^{-1} X^T e$$

and we see that $E\left[\lambda_1^T \hat{u}\right] = E\left[\lambda_2^T \hat{u}\right] = \alpha_1 - \alpha_2$. And

$$V\left[\lambda_1^T \hat{u}\right] = d_1 \sigma^2$$

$$V\left[\lambda_2^T \hat{u}\right] = 2\sigma_b^2 + d_2 \sigma^2$$

and

$$\mathrm{Cov}\left(\lambda_1^T\hat{u},\ \lambda_2^T\hat{u}\right) = \lambda_1^T A(X^TX)^{-1}A^T\lambda_2\sigma^2 = c\sigma^2$$

Let

$$\hat{\sigma}^2 = \frac{(Y - X\hat{u})^T(Y - X\hat{u})}{n - p + r}$$

It has been shown that $\hat{u}$ and hence $\lambda^T\hat{u}$ is independent of $\hat{\sigma}^2$. Now let Δ^T be any hypothesis such that $\Delta^T 1 = 0$ and $\Delta^T U_1 = 0$, then

$$E[SS(\Delta^T)] = \left[\sum_{j=1}^{s} \frac{\delta_j^T V_1 V_1^T \delta_j}{d_j}\right]\sigma_b^2 + s\sigma^2$$

Now

$$\hat{\sigma}_b^2 = \frac{SS(\Delta^T) - s\hat{\sigma}^2}{k}$$

where

$$k = \sum_{j=1}^{s} \frac{\delta_j^T V_1 V_1^T \delta_j}{d_j}$$

Now if $\lambda_i^T \varepsilon S$, then $\lambda_i^T\hat{u}$ and $\hat{\sigma}_b^2$ are uncorrelated. This can be seen from the following:

$$\mathrm{Cov}\left[\lambda_i^T\hat{u},\ \hat{\sigma}_b^2\right] = \mathrm{Cov}\left[\lambda_i^T\hat{u},\ \frac{\sum\left(\delta_j^T\hat{u}\right)^2\Big/ d_j - s\sigma^2}{k}\right]$$

$$\text{Cov}\left[\lambda_i^T\hat{u},\ \hat{\sigma}_b^2\right] = \sum_{i=1}^{s} \text{Cov}\left[\lambda_i^T\hat{u},\ \frac{\hat{u}^T\delta_j\delta_j^T\hat{u}}{d_j k}\right] - \frac{s}{k}\text{Cov}\left[\lambda_i^T\hat{u},\ \hat{\sigma}^2\right]$$

But $\text{Cov}\left[\lambda_i^T\hat{u},\ \hat{\sigma}^2\right] = 0$ and by Theorem 1.12

$$\text{Cov}\left[\lambda_i^T\hat{u},\ \hat{u}^T\delta_i\delta_i^T\hat{u}\right] = 2\lambda_i^T V\delta_i\delta_i^T u$$

where $u = E[\hat{u}] = \gamma 1 + U_i\alpha$. But $\delta_i^T u = 0$. Hence $\text{Cov}\left[\lambda_i^T\hat{u},\ \hat{u}^T\delta_i\delta_i^T\hat{u}\right] = 0$. Hence $\text{Cov}\left(\lambda_i^T\hat{u},\ \hat{\sigma}_b^2\right) = 0$. Now set

$$\hat{V} = \begin{bmatrix} d_1\hat{\sigma}^2 & c\hat{\sigma}^2 \\ c\hat{\sigma}^2 & 2\hat{\sigma}_b^2 + d_2\hat{\sigma}^2 \end{bmatrix}$$

and

$$a = \frac{\hat{V}^{-1}\ell}{\ell^T\hat{V}^{-1}\ell}$$

then

$$a^T\begin{bmatrix} \lambda_1^T\hat{u} \\ \lambda_2^T\hat{u} \end{bmatrix}$$

is an unbiased estimate of $\alpha_1-\alpha_2$. The choice of Δ^T and λ_1^T, λ_2^T and so on will not be discussed here. Let us note that this technique is used by many authors when analyzing the Balanced Incomplete Block Design when treatments are fixed and blocks are considered random. For this design, they usually choose λ_1^T and λ_2^T to represent the so-called interblock and intrablock estimates. Also they choose Δ^T so that Δ^T represents the hypothesis of treatments adjusted for blocks.

In other words, they use $t_1 = \hat{u}_{ij} - \hat{u}_{i'j}$ as one estimate of $\alpha_i - \alpha_{i'}$. The other estimate they use is $t_2 = \bar{u}'_{i.} - \bar{u}'_{i'.}$ where $\bar{u}'_{i.} = \sum_j n_{ij} y_{.j.}/(r - \lambda)$. Now due to the construction of the BIB design, $\bar{u}'_{i.} - \bar{u}'_{i'.}$ is independent of $\hat{u}_{ij} - \hat{u}_{i'j}$. The estimate of σ_b^2 is obtained by testing H_o: $u_{ij} = u_{ij'}$, and calculating the estimate of σ_b^2 in the usual manner. Their estimate of $\alpha_i - \alpha_{i'}$ is

$$t = a_1 t_1 + a_2 t_2$$

where

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \dfrac{\hat{V}^{-1}\ell}{\ell^T \hat{V}^{-1}\ell} \end{bmatrix}$$

and where $\hat{V}$ is a diagonal matrix with elements that are functions of $\hat{\sigma}^2$ and $\hat{\sigma}_b^2$. Note that this procedure is merely a special case of the procedure described above, that is the hypotheses λ_1^T and λ_2^T associated with $u_{ij} - u_{i'j}$ and $\bar{u}'_{i.} - \bar{u}'_{i'.}$ are elements of S and $\Delta^T 1 = 0$ and $\Delta^T U_1 \alpha = 0$. Since we are not restricting the estimates of $\alpha_i - \alpha_{i'}$ to be independent, it is possible to obtain other unbiased estimates than those usually suggested by most authors. A numerical example will be given in Chapter 6.

4.2.4 Point Estimation in the General Mixed Model

In the previous section, we restricted our discussion to the two-way classification without interaction. The methods described there are completely general and can be extended to other models. In this section, we shall consider a general mixed model and discuss another procedure for estimating $\lambda^T u$. Let $Y = Xu + e$ subject to $\theta^T u = 0$. Now suppose $e \sim N(0,V)$. If V is known, then the B.L.U.E. of u is given by

$$\bar{u} = [I - (X^T V^{-1} X)^{-1} \theta (\theta^T (X^T V^{-1} X)^{-1} \theta)^{-1} \theta^T][X^T V^{-1} X]^{-1} X^T V^{-1} Y$$

Hence the B.L.U.E. of $\lambda^T u$ is given by $\lambda^T \bar{u}$. In most cases V is not known. When this happens it is possible to replace V by $\hat{V}$, an estimate of V . If we want $\bar{u}$ to remain unbiased, we can choose $\hat{V}$ so that $\hat{V}$

and e are uncorrelated, that is suppose we replace V by $\hat{V}$, then

$$\bar{u}^* = [I - (X^T\hat{V}^{-1}X)^{-1}\theta(\theta^T(X^T\hat{V}^{-1}X)^{-1}\theta)^{-1}\theta^T]$$

$$[X^T\hat{V}^{-1}X]^{-1}X^T\hat{V}^{-1}[Xu + e]$$

or $\bar{u}^* = u + A^*e$ where

$$A^* = [I - (X^T\hat{V}^{-1}X)^{-1}\theta(\theta^T(X^T\hat{V}^{-1}X)^{-1}\theta)^{-1}\theta^T]$$

$$[X^T\hat{V}^{-1}X]^{-1}X^T\hat{V}^{-1}$$

Now if $\hat{V}$ and e are uncorrelated, then

$$E[A^*e] = E[A^*]E[e] = E[A^*] \cdot 0 = 0$$

Thus $E[\bar{u}^*] = u$. An example of this approach is given in Chapter 6.

CHAPTER 5. OTHER TOPICS

5.0 Preliminaries

In this chapter, we shall consider other topics related to the analysis of linear models. These topics include regression, covariance, and a discussion of tests for main effects and interactions. A section on the construction of θ^T also is included.

5.1 Regression

The "u" model as defined in Chapter 2 is the classical design model. We could have started with the general regression model and then observed that the "u" model is a special case; however, since we wanted to emphasize the use of the "u" model in experimental design, we chose to start with the "u" model as defined in Chapter 2. We shall now define the general regression model.

Let $e \sim N(0,\sigma^2 I)$ and let $Y = Xu + e$ subject to $\theta^T u = 0$. We no longer consider u_{ij} as the mean of the $(ij)th$ population and $X^T X$ is not generally diagonal. Likewise, $u^* = (X^T X)^{-1} X^T Y$ is no longer the "cell" means. The theory developed in Chapter 2 is completely general, and by changing the interpretation of u, $\hat{u}$, u^*, and X, we can use the results developed there. The classical regression equation is

$$Y = Xu + e$$

Usually, there are no restrictions placed on the parameters. Hence we see that $\hat{u} = (X^TX)^{-1}X^TY$. If H_o: $\Lambda^T u = \xi$ then

$$F = \frac{[\Lambda^T\hat{u} - \xi]^T(\Lambda^T(X^TX)^{-1}\Lambda)^{-1}[\Lambda^T\hat{u} - \xi]}{s\hat{\sigma}^2}$$

where

$$\hat{\sigma}^2 = \frac{Y^T[I - X(X^TX)^{-1}X^T]Y}{n - p + r}$$

These are all well-known results.

If there are restrictions placed on the parameters, that is $\theta^T \neq 0$, then $\hat{u}$, $\hat{\sigma}^2$ and F can be found by the techniques described in Chapter 2. Since the analysis of these models is well known, they will not be discussed further.

5.2 Covariance Model

In this section, we will consider a special case of the regression model, that is

$$y_{ijk} = u_{ij} + \sum_{p=1}^{m} \beta_p z_{ijkp} + e_{ijk}$$

subject to $\theta^T u = 0$. Or in matrix notation, we have

$$Y = Xu + Z\beta + e \tag{5.1}$$

subject to $\theta^T u = 0$ where

X is the $n \times p$ design matrix of rank p

u is the vector of u_{ij}

Z is the $n \times m$ matrix associated with β

β is the vector of covariates

$e \sim N(0, \sigma^2 I)$.

Let us consider the density of Y

$$f(y;u,\beta,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \mathrm{EXP}\left\{-\frac{(Y - Xu - Z\beta)^T(Y - Xu - Z\beta)}{2\sigma^2}\right\}$$

subject to $\theta^T u = 0$. We will find maximum likelihood estimates of u, β, σ^2 . As is customary, we maximize

$$\ln f(y;u,\beta,\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{(Y - Xu - Z\beta)^T(Y - Xu - Z\beta)}{2\sigma^2}$$

subject to $\theta^T u = 0$. The Lagrangian function is

$$L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{(Y - Xu - Z\beta)^T(Y - Xu - Z\beta)}{2\sigma^2} + \frac{2\delta^T\theta^T u}{\sigma^2}$$

$$\frac{\partial L}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{(Y - Xu - Z\beta)^T(Y - Xu - Z\beta)}{2\sigma^4} - \frac{2\delta^T\theta^T u}{\sigma^4} = 0 \tag{1}$$

$$\frac{\partial L}{\partial u} = \frac{2(-X^TY + (X^TX)u + X^TZ\beta + \theta\delta)}{2\sigma^2} = 0 \tag{2}$$

$$\frac{\partial L}{\partial\beta} = \frac{2(-Z^TY + (Z^TZ)\beta + Z^TXu)}{2\sigma^2} = 0 \tag{3}$$

$$\frac{\partial L}{\partial\delta^T} = \frac{2\theta^T u}{\sigma^2} = 0 \tag{4}$$

Now from (1) and (4) we get

$$\tilde{\sigma}^2 = \frac{[Y - X\hat{u} - Z\beta]^T[Y - X\hat{u} - Z\beta]}{n}$$

From (2), we see that

$$\hat{u} = (X^TX)^{-1}X^TY - (X^TX)^{-1}[X^TZ\hat{\beta} + \theta\hat{\delta}]$$

But from (4) we see that $\theta^T\hat{u} = 0$ hence

$$0 = \theta^T\hat{u} = \theta^T(X^TX)^{-1}X^TY - \theta^T(X^TX)^{-1}X^TZ\hat{\beta} - (\theta^T(X^TX)^{-1}\theta)\hat{\delta}$$

or

$$\hat{\delta} = (\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T[(X^TX)^{-1}X^TY - (X^TX)^{-1}X^TZ\hat{\beta}]$$

Hence

$$\begin{aligned}\hat{u} &= (X^TX)^{-1}X^TY - (X^TX)^{-1}[X^TZ\hat{\beta}] \\ &\quad - (X^TX)^{-1}\theta[\theta^T(X^TX)^{-1}\theta]^{-1}\theta^T(X^TX)^{-1}X^TY \\ &\quad - (X^TX)^{-1}\theta[\theta^T(X^TX)^{-1}\theta]^{-1}\theta^T(X^TX)^{-1}X^TZ\hat{\beta} \\ &= [I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]u^* \\ &\quad - [I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]\gamma^*\end{aligned}$$

where

$$u^* = (X^TX)^{-1}X^TY$$

$$\gamma^* = (X^TX)^{-1}X^TZ\hat{\beta}$$

Let

$$A = [I - (X^TX)^{-1}\theta(\theta^T(X^TX)^{-1}\theta)^{-1}\theta^T]$$

then

$$\hat{u} = A[u^* - \gamma^*] \tag{5.2}$$

Now from (3), we see that

$$(Z^TZ)\hat{\beta} = Z^TY - Z^TX\hat{u}$$

$$= Z^TY - Z^TXA[u^* - \gamma^*]$$

$$= Z^TY - Z^TXAu^* + Z^TXA(X^TX)^{-1}X^TZ\hat{\beta}$$

or

$$[Z^TZ - Z^TXA(X^TX)^{-1}X^TZ]\hat{\beta} = Z^TY - Z^TXA(X^TX)^{-1}X^TY$$

$$= Z^T[I - XA(X^TX)^{-1}X^T]Y$$

or

$$Z^T[I - XA(X^TX)^{-1}X^T]Z\hat{\beta} = Z^T[I - XA(X^TX)^{-1}X^T]Y$$

or

$$\hat{\beta} = [Z^T(I - XA(X^TX)^{-1}X^T)Z]^{-1}Z^T[I - XA(X^TX)^{-1}X^T]Y \tag{5.3}$$

Now let

$$\hat{\sigma}^2 = \frac{[Y - X\hat{u} - Z\hat{\beta}]^T[Y - X\hat{u} - Z\hat{\beta}]}{n - p + r} \tag{5.4}$$

By following similar arguments as in Chapter 2, we can show that $\hat{u}$, $\hat{\beta}$, $\hat{\sigma}^2$ are minimum variance unbiased estimates. Now suppose we wish to test H_o: $\Delta^T\beta = 0$. Then

$$SS(\Delta^T) = (\Delta^T\hat{\beta})^T\left[\frac{V[\Delta^T\hat{\beta}]}{\sigma^2}\right]^{-1}(\Delta^T\hat{\beta}) \ . \qquad (5.5)$$

Now

$$V(\Delta^T\hat{\beta}) = \Delta^T V(\hat{\beta})\Delta$$

and

$$V(\hat{\beta}) = [Z^T(I - XA(X^TX)^{-1}X^T)Z]^{-1}\sigma^2$$

Now let us consider $V(\hat{u})$.

$$V(\hat{u}) = A[V(u^*) + V(\gamma^*) - 2\ \mathrm{Cov}\ (u^*,\gamma^*)]A^T$$

Now

$$V(u^*) = \sigma^2(X^TX)^{-1}$$

$$V(\gamma^*) = (X^TX)^{-1}X^TZV(\hat{\beta})Z^TX(X^TX)^{-1}$$

$$\begin{aligned}
\mathrm{Cov}\ (u^*,\gamma^*) &= \mathrm{Cov}\ [(X^TX)^{-1}X^TY,\ (X^TX)^{-1}X^TZ\hat{\beta}] \\
&= (X^TX)^{-1}X^T[\mathrm{Cov}\ (Y,\hat{\beta})]Z^TX(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T[I - X(X^TX)^{-1}A^TX^T] \\
&\quad \cdot ZV(\hat{\beta})Z^T(X^TX)^{-1} \\
&= [(X^TX)^{-1}X^T - (X^TX)^{-1}A^TX^T]ZV(\hat{\beta}) \\
&\quad \cdot Z^T(X^TX)^{-1}
\end{aligned}$$

But

$$A[\mathrm{Cov}\ (u^*,\gamma^*)]A^T = [A(X^TX)^{-1}X^T - A(X^TX)^{-1}A^TX^T] \cdot ZV(\hat{\beta})Z^T(X^TX)^{-1}A^T$$

$$= [A(X^TX)^{-1}X^T - A(X^TX)^{-1}X^T]ZV(\hat{\beta}) \cdot Z^T(X^TX)^{-1}A^T$$

$$= 0$$

Thus

$$V(\hat{u}) = A(X^TX)^{-1}\sigma^2 + A(X^TX)^{-1}X^TZV(\hat{\beta})Z^TX(X^TX)^{-1}A^T\sigma^2$$

Now suppose we wish to test H_o: $\Lambda^Tu = 0$, then

$$SS(\Lambda^T) = (\Lambda^T\hat{u})^T\left[\frac{\Lambda^T(V(\hat{u}))^{-1}\Lambda}{\sigma^2}\right]^{-1}(\Lambda^T\hat{u}) \tag{5.6}$$

Therefore Eq's. (5.5) and (5.6) provide us with direct tests of hypothesis for testing H_o: $\Lambda^T\beta = 0$ and H_o: $\Lambda^Tu = 0$ respectively.

Let us consider the following example taken from Ostle [24].

Gains in Weight (Y) and Initial Weights (X) of Pigs in a Feeding Trial

Treatment

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	30	165	24	180	34	156	41	201
	27	170	31	169	32	189	32	173
	20	130	20	171	35	138	30	200
	21	156	26	161	35	190	35	193
	33	167	20	180	30	160	28	142
	29	151	25	170	29	172	36	189
Total	160	939	146	1031	195	1005	202	1098

The model is

$$y_{ij} = u_i + \beta z_{ij} + e_{ij}$$

Now $A = I$, $(X^TX) = 6I$, $Z^TX = [160 \quad 146 \quad 195 \quad 202]$

Hence

$$Z^T[I - XA(X^TX)^{-1}X^T]Z = Z^TZ - \frac{(Z^TX)(X^TZ)}{6}$$

$$= 21319 - 20957.5 = 361.5$$

and

$$Z^T[I - XA(X^TX)^{-1}X^T]Y = Z^TY - \frac{(Z^TX)(X^TY)}{6}$$

$$120253 - 119756.17$$

$$= 496.83$$

Thus $\hat{\beta} = \frac{496.83}{361.5} = 1.374$ and $V(\hat{\beta}) = \sigma^2(361.5)^{-1}$.

Therefore, if we test H_o: $\beta = 0$, then

$SS(\beta) = 361.5(1.374)^2 = 682.5$.

Now

$$\hat{u} = A(u^* - \gamma^*)$$

$$= \overline{Y} - \frac{1}{6} X^TZ\hat{\beta}$$

$$\hat{u} = \begin{bmatrix} 156.50 \\ 171.83 \\ 167.50 \\ 183.00 \end{bmatrix} - \frac{1}{6}\begin{bmatrix} 160 \\ 146 \\ 195 \\ 202 \end{bmatrix}(1.374)$$

or

$$\hat{u} = \begin{bmatrix} 119.85 \\ 138.39 \\ 122.83 \\ 136.73 \end{bmatrix}$$

Suppose we wished to test H_o: $u_1 = u_2 = u_3 = u_4$ or H_o: $\Lambda^T u = 0$ where

$$\Lambda^T = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

Now

$$V(\hat{u}) = \begin{bmatrix} 2.13 & 1.79 & 2.40 & 2.48 \\ 1.79 & 1.81 & 2.19 & 2.27 \\ 2.40 & 2.19 & 3.09 & 3.03 \\ 2.48 & 2.27 & 3.03 & 3.30 \end{bmatrix} \sigma^2$$

Hence

$$SS(\Lambda^T) = (\Lambda^T \hat{u})^T \left[\frac{\Lambda^T V(\hat{u}) \Lambda}{\sigma^2}\right]^{-1} [\Lambda^T \hat{u}] = 1609.6$$

This concludes the discussion of covariance. Eq's. (5.2), (5.3), (5.4), (5.5), (5.6) provide the information necessary to analyze a general covariance model as described by Eq. (5.1).

5.3 The Construction of θ^T

In this section, we consider two methods of constructing θ^T . The first uses Theorem 3.0, that is, if we are given the "β" model $Y = XP\beta + e$, then we choose θ^T so that

(1) $\theta^T P = 0$

(2) $\text{rank}(\theta^T) = \text{rank}(X) - \text{rank}(XP)$.

This can be done by finding the linearly independent rows of $I-PP^+$. However, this requires computing P^+, which can be time consuming if P is a "large" matrix.

The second approach can be used for most of the classical experimental design models. Here we must know the functional form of $\theta^T u$. For example, in the two-way without interaction we know that the u_{ij} must satisfy

$$u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0$$

for all i, i', j, j'.

We will now give rules for finding θ^T for the two-way classification without interaction, and the three-way classification with interaction. First, we will assume no missing cells and then consider the case when there are missing cells.

5.3.1 The Two-Way Classification Without Interaction

Let us consider the following model.

$$y_{ijk} = u_{ij} + e_{ijk}$$

$i = 1, 2, 3, 4, 5$; $j = 1, 2, 3, 4$; $k = 1,\cdots,n_{ij}$ subject to $u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0$ for all i, i', j, j'.

We can generate θ^T in the following manner.

(1) Write $u_{11} - u_{I1} - u_{1J} + u_{IJ}$

(2) Let J vary from 2 to 4 with $I = 2$.

(3) Let J vary from 2 to 4 with $I = 3$.

(4) Let J vary from 2 to 4 with $I = 4$.

(5) Let J vary from 2 to 4 with $I = 5$.

In computer notation, we would have

```
      L1  =  1
      L2  =  1
      DO 1  I  =  2,5
      DO 1  J  =  2,4
      WRITE (6,30) L1,L2,I,L1,L2,J,I,J
    1 CONTINUE
   30 FORMAT (1H , 4(2I3,2X))
```

This would print out the indices of the u_{ij}'s. For instance, the output would look like

11	21	12	22
11	21	13	23
11	21	14	24
11	31	12	32
11	31	13	33
11	31	14	34
11	41	12	42
11	41	13	43
11	41	14	44
11	51	12	52
11	51	13	53
11	51	14	54

Thus, the first row of $\theta^T u$ is

$$u_{11} - u_{21} - u_{12} + u_{22} = 0 ;$$

the second row is $u_{11} - u_{21} - u_{13} + u_{23} = 0$, and so on. This is a linearly independent set such that $\text{rank}(\theta^T) = 15$ as it should.

5.3.2 The Three-Way Classification Without Interaction

Suppose our model is

$$y_{ijk\ell} = u_{ijk} + e_{ijk\ell}$$

$i = 1,\cdots,a$; $j = 1,\cdots,b$; $k = 1,\cdots,e$; $\ell = 1,\cdots,n_{ijk}$ subject to no two-way interactions and no three-way interactions. This means

$$u_{ij\cdot} - u_{i'j\cdot} - u_{ij'\cdot} + u_{i'j'\cdot} = 0 \tag{1}$$

$$u_{\cdot jk} - u_{\cdot j'k} - u_{\cdot jk'} + u_{\cdot j'k'} = 0 \tag{2}$$

$$u_{i\cdot k} - u_{i'\cdot k} - u_{i\cdot k'} + u_{i'\cdot k'} = 0 \tag{3}$$

$$\begin{aligned} u_{ijk} - u_{i'jk} - u_{ij'k} + u_{i'j'k} \\ - u_{ijk'} + u_{i'jk'} + u_{ij'k'} \\ - u_{i'j'k'} = 0 \end{aligned} \tag{4}$$

for all i, i', j, j', k, k' .

To generate a linearly independent set satisfying (1), we use

$$u_{11\cdot} - u_{I1\cdot} - u_{1J\cdot} + u_{IJ\cdot} = 0$$

and let $I = 2,\cdots,a$, and $J = 2,\cdots,b$ in the same manner as in Section 5.3.1. A set satisfying (2) can be obtained by using

$$u_{\cdot 11} - u_{\cdot J1} - u_{\cdot 1K} + u_{\cdot JK} = 0$$

and letting $J = 2,\cdots,b$; $K = 2,\cdots,c$.

For (3) we use

$$u_{1\cdot 1} - u_{I\cdot 1} - u_{1\cdot K} + u_{I\cdot K} = 0$$

As $I = 2,\cdots,a$ and $K = 2,\cdots,c$. For a set satisfying (4), we use

$$u_{111} - u_{I11} - u_{1J1} + u_{IJ1} - u_{11K} + u_{I1K} + u_{1JK} - u_{IJK}$$

for $I = 2,\cdots,a$; $J = 2,\cdots,b$; $K = 2,\cdots,c$. We must fix I and J and then let K vary. In computer notation, we have a nested "do-loop."

For example, suppose

$$I = 1, 2, 3; \quad J = 1, 2, 3; \quad K = 1, 2, 3 .$$

The set representing no (I,J) interaction is

$$u_{11\cdot} - u_{21\cdot} - u_{12\cdot} + u_{22\cdot} = 0$$

$$u_{11\cdot} - u_{21\cdot} - u_{13\cdot} + u_{23\cdot} = 0$$

$$u_{11\cdot} - u_{31\cdot} - u_{12\cdot} + u_{32\cdot} = 0$$

$$u_{11\cdot} - u_{31\cdot} - u_{13\cdot} + u_{33\cdot} = 0$$

The sets representing no (J,K) and no (I,K) interactions can be obtained in an analogous manner.

The set representing no (I,J,K) interaction is

$$
\begin{aligned}
&u_{111} - u_{221} - u_{121} + u_{221} \\
&- u_{112} + u_{212} + u_{122} - u_{222} \quad = \quad 0
\end{aligned}
$$

$$
\begin{aligned}
&u_{111} - u_{211} - u_{121} + u_{221} \\
&- u_{113} - u_{213} - u_{123} - u_{223} \quad = \quad 0
\end{aligned}
$$

$$
\begin{aligned}
&u_{111} - u_{211} - u_{131} + u_{231} \\
&- u_{112} + u_{212} + u_{132} - u_{232} \quad = \quad 0
\end{aligned}
$$

$$
\begin{aligned}
&u_{111} - u_{211} - u_{131} + u_{231} \\
&- u_{113} + u_{213} + u_{133} - u_{233} \quad = \quad 0
\end{aligned}
$$

$$
\begin{aligned}
&u_{111} - u_{311} - u_{121} + u_{321} \\
&- u_{112} + u_{312} + u_{122} - u_{322} \quad = \quad 0
\end{aligned}
$$

$$
\begin{aligned}
&u_{111} - u_{311} - u_{121} + u_{321} \\
&- u_{113} + u_{313} + u_{123} - u_{323} \quad = \quad 0
\end{aligned}
$$

$$
\begin{aligned}
&u_{111} - u_{311} - u_{131} + u_{331} \\
&- u_{112} + u_{312} + u_{132} - u_{332} \quad = \quad 0
\end{aligned}
$$

$$
\begin{aligned}
&u_{111} - u_{311} - u_{131} + u_{331} \\
&- u_{113} + u_{313} + u_{133} - u_{333} \quad = \quad 0
\end{aligned}
$$

Thus, for this example θ^T is a (20×27) matrix of rank 20.

5.3.3 Split Plot

As shown in Chapter 3, we can consider this as a three-way classification with no (J,K) and no (I,J,K) interaction. Thus the model would be

$$y_{ijk} = u_{ijk} + e_{ijk}$$

subject to

$$u_{\cdot jk} - u_{\cdot j'k} - u_{\cdot jk'} + u_{\cdot j'k'} = 0$$

$$\begin{aligned} u_{ijk} - u_{i'jk} - u_{ij'k} + u_{i'j'k} \\ - u_{ijk'} + u_{i'jk'} + u_{ij'k'} \\ - u_{i'j'k'} &= 0 \end{aligned}$$

for all i, i', j, j', k, k' . We construct θ^T by finding the linearly independent set for no (J,K) interaction and the set for no (I,J,K) interaction.

5.3.4 Missing Cells

The procedure for constructing θ^T when there are missing cells is as follows.

(1) Construct θ^T as though no cells were missing.

(2) Eliminate the u_{ij}'s that do not exist in the model.

Consider this example; $y_{ij} = u_{ij} + e_{ij}$, where $i = 1,2,3$; $j = 1,2,3$, and $k = 0,1,\cdots,n_{ij}$

i \ j	1	2	3
1	X		X
2	X	X	X
3	X	X	X

In other words, u_{12} is not in our model. Suppose the constraints are no (I,J) interaction. Now we write

$$u_{11} - u_{I1} - u_{1J} + u_{IJ} = 0$$

and generate

$$u_{11} - u_{21} - u_{12} + u_{22} = 0 \qquad (1)$$

$$u_{11} - u_{21} - u_{13} + u_{23} = 0 \qquad (2)$$

$$u_{11} - u_{31} - u_{12} + u_{32} = 0 \qquad (3)$$

$$u_{11} - u_{31} - u_{13} + u_{33} = 0 \qquad (4)$$

Now we subtract (3) from (1) and obtain

$$u_{31} - u_{32} - u_{21} + u_{22} = 0$$

Now

$$u_{31} - u_{21} - u_{32} + u_{22} = 0$$

$$u_{11} - u_{21} - u_{13} + u_{23} = 0$$

$$u_{11} - u_{31} - u_{13} + u_{33} = 0$$

are a set of linearly independent constraints which span the set of possible constraints. Hence we take $\theta^T u$ to be the above three equations.

5.3.5 Balanced Incomplete Block

As shown in Chapter 3, we can consider this as a two-way classification without interaction and with missing cells. For example: $y_{ij} = u_{ij} + e_{ij}$ subject to no (I,J) interaction. Suppose we observed

i \ j	1	2	3
1	X	X	
2	X		X
3		X	X

that is u_{31}, u_{22}, and u_{13} are not in the model.

Again we write the constraints as though there were no missing cells, that is

$$u_{11} - u_{21} - u_{12} + u_{22} = 0 \qquad (1)$$

$$u_{11} - u_{21} - u_{13} + u_{23} = 0 \qquad (2)$$

$$u_{11} - u_{31} - u_{12} + u_{32} = 0 \qquad (3)$$

$$u_{11} - u_{31} - u_{13} + u_{33} = 0 \qquad (4)$$

We must eliminate u_{31}, u_{22}, and u_{13} from the above equations. We subtract (4) from (3) and obtain

$$u_{13} - u_{12} - u_{33} + u_{32} = 0 \qquad (5)$$

Subtracting (2) from (1) we obtain

$$-u_{12} + u_{13} - u_{23} + u_{22} = 0 \qquad (6)$$

Subtracting (6) from (5) we have

$$-u_{33} + u_{32} + u_{23} - u_{22} = 0$$

or

$$u_{22} = -u_{33} + u_{32} + u_{23}$$

Substituting u_{22} into (1) we have

$$u_{11} - u_{21} - u_{12} + u_{32} - u_{33} + u_{23} = 0$$

Thus

$$\theta^T u = u_{11} - u_{21} - u_{12} + u_{32} - u_{33} + u_{23}$$

5.3.6 Latin Square

The Latin Square can be considered as a three-way classification with no interactions and with missing cells.

Suppose we observe

		Column 1	Column 2	Column 3
Row	1	y_{111}	y_{122}	y_{133}
	2	y_{212}	y_{223}	y_{231}
	3	y_{313}	y_{321}	y_{332}

Here we have a 3×3 Latin Square. We can consider the model as $y_{ijk} = u_{ijk} + e_{ijk}$ subject to no interactions where $i = 1,2,3$; $j = 1,2,3$; and $k = 0,1,2,n_{ij}$. Let us first write θ^T if we had no missing cells. Now θ^T can be constructed by using the results of Section 5.3.2. We would then eliminate all the u_{ijk}'s except $u_{111}, u_{122}, u_{133}, u_{212}, u_{223}, u_{231}, u_{313}, u_{321}, u_{332}$. The result of this is

$$\theta^T u = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 & -1 & 0 & 1 \\ 1 & 0 & -1 & -1 & 1 & 0 & 0 & -1 & 1 \end{bmatrix} u$$

Naturally we would not use this approach if we had no missing cells since standard computing formulas are readily available.

5.4 Tests of Hypothesis

Previously, we have assumed we had the hypothesis to be tested. There was no discussion about whether these hypotheses were appropriate. Our attitude has been that if an experimenter came to us with a model and wanted to test a given hypothesis, say H_o: $\Lambda^T u = 0$, then we provided the correct test. It was the experimenter's problem to decide what he wanted to test. This attitude seems to contradict most textbooks on applied statistics and most computer programs that analyze linear models.

These textbooks and programs provide the reader or user with one set of hypotheses that are to be tested. They seem to say, "Here are the answers, we hope you have asked the questions that correspond to these answers."

Let us take a simple example. Suppose an experimenter came to us with data taken in a completely randomized design in which he had three "treatments." He knew before he ran the experiment that the effects due to the "treatments" were not equal. What he wanted to know was, "Is the effect of treatment 1 twice the effect of treatment 2 plus treatment 3?"

Now he runs his experiment, obtains the data, and calls on the local statistician to analyze his data. The statistician is overjoyed at having such a simple textbook problem. In no time he gives the experimenter the usual computer output:

Source of Variation	df	SS	MS	F
"Treatments"	a-1	XX	XX	f**
Error	n-a	XX	XX	

As the statistician hands the output to the experimenter, he happily explains, "Your F value is highly significant." The experimenter, seeing the statistician is pleased about the significant F, also is pleased and leaves with great admiration for the mystic powers of the statistician and his science.

Farfetched? Not at all; similar occurrences happen daily. The statistician is given data and asked to analyze it. He puts the data into one of the "can" programs, gets the standard output, and thinks he has done his job. This is fine as long as he is in the simple balanced case. But things change when he tries to analyze data in the unbalanced case. In the first place, the literature on analyzing such data is quite confusing. This problem stems, possibly, from trying to tie the analysis of linear models to the analysis of "main" effects and interactions.

For example, suppose we are given the following table:

Source of Variation

A	a-1
B	b-1
Error	(a-1) (b-1)

It is customary to refer to A and B as main effects. There would be general agreement on the interpretation given to them. If we are given:

Source of Variation

A	(a-1)
B	(b-1)
AB	(a-1) (b-1)
Error	ab(r-1)

we still say A and B are main effects and AB is interaction. Now the trouble begins. A number of texts indicate that "A" means that we are testing H_o: α: = 0, i = 1,$\cdots$,a . They give this interpretation to both of the previously given tables. But then we read,

> "In the presence of interaction, it is rarely useful to ask about main effects of either factor.....Despite questionable meaning, mean squares for the main effects are commonly reported even when interaction is present. This is a routine practice which is open to criticism, but is almost unavoidable."

One text defines the main effect A "as a measure of the change in the response variable to changes in the level of the factor averaged over all levels of all the other factors." But in the two-way model with interaction, this same text indicates the test for A implies testing H_o: $\alpha_1 = \cdots = \alpha_a$. There is no averaging here. Also the fact that we are in a balanced case allows us to do things that are not valid in the unbalanced case. For instance, consider the following two tables:

Source of Variation	df	SS
A	a-1	SSA
B	(b-1)	SSB
AB	(a-1) (b-1)	SSAB
Error	ab(n-1)	E

and

Source of Variation	df	SS
A	a-1	SSA
B	b-1	SSB
Error	abn-a-b+1	E'

Note that one difference in the two tables is that $E' = E + SSAB$ and that SSA and SSB remain the same. If we have unbalanced data, this is not true. We would get an SSA' and SSB' as well. In the unbalanced case, what does SSA, SSB, SSA' and SSB' signify?

Suppose we are in the "missing" cell case and the model is $y_{ijk} = u + \alpha_i + \beta_j + \delta_{ij} + e_{ijk}$. One textbook says "it is now necessary to assume that there are no empty cells, else the main effects are not estimable..." Yet another text presents methods to obtain the sum for "main" effects in this case.

This brings us to another point. Are the restrictions placed on the parameters (or random variables) an integral part of the model or are they used simply to help solve the normal equations? Again, if we are in the balanced case, it really makes little difference. But what about the unbalanced case! The use of $\sum \alpha_i = 0$ rather than $\sum n_{ij}\alpha_i = 0$ can result in some very confusing answers in the analysis of the two-way interaction design. Also, if we are in the mixed model, the expected values of the mean squares depend upon whether we use the restrictions as an integral part of the model.

Summing up, it seems that an analysis has been developed for certain models in the balanced case. This analysis was tied to the idea of main effects and interactions. When other models were developed, their analyses were forced to flow along lines similar to the balanced models. Since the techniques, concepts and interpretations were special to the balanced case, they could not be carried over directly to the unbalanced case. There is no need to tie the analysis to one form. In fact, there is no reason to think that experimenter A and experimenter B want the same analysis to

$$y_{ijk} = u + \alpha_i + \gamma_j + \delta_{ij} + e_{ijk} .$$

It is absurd to think both men are interested in the same hypotheses. Their analyses should depend upon the reasons for which they conducted their experiments. Their experiments should have been designed to answer certain questions.

Does this "u"-model approach free the statistician from all responsibility in formulating the hypothesis to be tested? Yes and no. The statistician, in his role as consultant, will have to take the objectives of his client and express them in meaningful mathematical terms. As we all know, the mere statement of the objectives is quite an accomplishment. Also, in certain models, the statistician can present hypotheses suggested by the model. For example, suppose we had a 3×3 Latin Square with one missing observation. Let us suppose u_{111} is missing. And suppose we would like to suggest some possible hypotheses to be tested.

	1	2	3	
1	A	B u_{122}	C u_{133}	$u_{1\cdot\cdot}$
2	B u_{212}	C u_{223}	A u_{231}	$u_{2\cdot\cdot}$
3	C u_{313}	A u_{321}	B u_{332}	$u_{3\cdot\cdot}$
	$u_{\cdot1\cdot}$	$u_{\cdot2\cdot}$	$u_{\cdot3\cdot}$	

Some possible hypotheses are

$$(1) \quad H_o: \ \frac{u_{1\cdot\cdot}}{2} = \frac{u_{2\cdot\cdot}}{3} = \frac{u_{3\cdot\cdot}}{3}$$

$$(2) \quad H_o: \ \frac{u_{\cdot 1\cdot}}{2} = \frac{u_{\cdot 2\cdot}}{3} = \frac{u_{\cdot 3\cdot}}{3}$$

(3) H_o: $K_1\bar{u}_{1\cdot\cdot} + K_2\bar{u}_{2\cdot\cdot} + K_3\bar{u}_{3\cdot\cdot} = 0$ where

$\sum K_i$ is not necessarily zero.

and so on.

Another example is the following two-way classification with interaction

	1	2	3
1	u_{11}		u_{13}
2	u_{21}	u_{22}	u_{23}
3		u_{32}	u_{33}
4	u_{41}	u_{42}	

What are appropriate hypotheses for this model? The first, of course, should be to see if the model is additive. Suppose not, then what can we ask. We could test the following:

$$H_o: \ u_{11} = u_{21} = u_{41}$$

$$H_o: \ u_{22} = u_{32} = u_{42}$$

$$H_o: \ u_{13} = u_{23} = u_{33}$$

Or we could use a test like the one suggested by Elston and Bush [12]. Again, we should define the purpose of the experiment and ask questions with that in mind.

In conclusion, let us note that throughout Chapter 3 we postponed the discussion of the "merits" of various hypotheses to this section. Yet nowhere in this section did we discuss the "merits" of these hypotheses. The reason for this is that, while any of the hypotheses that were posed are legitimate, their value or merit depends upon the purpose of the experiment. The researcher and the statistician have a joint responsibility to pose meaningful hypotheses that are to be tested. Finally it might be better if we did not tie hypotheses to the idea of "main" effects and interactions.

CHAPTER 6. NUMERICAL EXAMPLES

6.0 Preliminaries

We will now consider several numerical examples. These examples were run on the UNIVAC 1108 using a program written in FORTRAN V. The program calculates the sum of squares for H_o: $\lambda^T u = 0$ in the following manner.

(1) λ^T is orthogonalized, that is $\lambda_1^T = H_1 \lambda^T$ where H_1 is chosen so that
$\lambda_1^T \lambda_1 = H_1 \lambda^T \lambda H_1^T = I$

(2) $\lambda_2^T = H_2 \lambda_1^T$ where H_2 is chosen so that

$$\lambda_2^T A(X^T X)^{-1} A^T \lambda_2 = H_2 \lambda_1^T A(X^T X)^{-1} A^T \lambda_1 H_2^T = D$$

(3) $$SS(\lambda^T) = S(\lambda_2^T) = \sum_{i=1}^{a} \frac{1}{d_i} (\lambda_{2i}^T \hat{u})^2$$

This procedure is explained in detail in Chapter 4. Because of a lack of space, this chapter does not consider all the intermediate steps needed to calculate the sum of squares.

6.1 The One-Way Classification

Consider the following example taken from Ostle [24].

Storage Conditions

1	2	3	4	5
7.3	5.4	8.1	7.9	7.1
8.3	7.4	6.4	9.5	
7.6	7.1		10.0	
8.4				
8.3				

The model is

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

where

$$u_i = u + \alpha_i$$

The hypothesis to be tested is

$$H_o: \ u_1 = u_2 = u_3 = u_4 = u_5$$

which is equivalent to H_o: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$ in the "β" model.

Thus $\Lambda_1^T u = 0$ is

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix} u = 0$$

Let

$$\Lambda^T = \begin{bmatrix} .8617 & -.3463 & -.1258 & -.3463 & -.0432 \\ .1713 & .3691 & -.8321 & .3690 & -.0773 \\ 0 & .7071 & 0 & -.7071 & 0 \\ .1679 & .2095 & .3031 & .2095 & -.8900 \end{bmatrix}$$

$$D = \begin{bmatrix} .2382 & 0 & 0 & 0 \\ 0 & .4488 & 0 & 0 \\ 0 & .0 & .3333 & 0 \\ 0 & 0 & 0 & .8730 \end{bmatrix}$$

Then H_o: $\Lambda^T u = 0$ is equivalent to H_o: $\Lambda_1^T u = 0$ where $\Lambda^T \Lambda = I$ and $\Lambda^T (X^T X)^{-1} \Lambda = D$.

Now $\hat{u} = u^* = (X^T X)^{-1} X^T Y$.

But $(X^T X)$ is diagonal, hence $\hat{u}$ is just the cell means, that is

$$\hat{u} = \begin{bmatrix} 8.0 \\ 6.6 \\ 7.3 \\ 9.1 \\ 7.1 \end{bmatrix}$$

and

$$SS(\Lambda^T) = \sum_{i=1}^{4} \frac{1}{d_i} \left(\lambda_i^T \hat{u}\right)^2 = 10.66$$

and

$$\hat{\sigma}^2 = \frac{[Y - X\hat{u}]^T [Y - X\hat{u}]}{14 - 5} = \frac{7.17}{9} = .80$$

and

$$F = \frac{10.66}{4(.80)} = 3.34$$

Now suppose $\alpha_i \sim N\left(0, \sigma_a^2\right)$, and we wish to estimate σ_a^2 using the above $SS(\Lambda^T)$. From Eq. 4.7, we see that

$$\hat{\sigma}_a^2 = \frac{SS(\Lambda^T) - 4\hat{\sigma}^2}{k_0}$$

where $k_0 = \sum \frac{1}{d_i} = 10.56$

Hence

$$\hat{\sigma}_a^2 = \frac{10.66 - 4(.80)}{10.56} = .7064$$

and the $V\left(\hat{\sigma}_a^2\right)$ is, from Eq. (4.8),

$$V\left(\hat{\sigma}_a^2\right) = \frac{2k_1^2}{k_0^2}\sigma_a^4 + \frac{4}{k_0}\sigma_a^2\sigma^2 + \frac{2}{k_0^2}\left[4 + \frac{16}{9}\right]\sigma^4$$

where $k_1 = \sum \frac{1}{d_i^2} = 32.898$ or

$$V\left(\hat{\sigma}_a^2\right) = .59\sigma_a^4 + .38\sigma_a^2\sigma^2 + .104\sigma^4$$

which agrees with the results obtained from the formula in Searle's paper [29].

Now suppose we wanted to estimate σ_a^2 from H_o: $\Lambda^T u = 0$ where

$$\Lambda^T = \frac{1}{\sqrt{2}}[0 \quad 1 \quad 0 \quad -1 \quad 0]$$

Then we see that $D = .3333$ and

$$SS(\Lambda^T) = \frac{1}{.3333}(\lambda^T\hat{u})^2 = 3\left(\frac{2.5}{\sqrt{2}}\right)^2 = \frac{3}{2}(6.25)$$

$$= 9.375$$

Thus

$$\tilde{\sigma}_a^2 = \frac{9.375 - 3.20}{k_0}$$

where

$$k_0 = \sum \frac{1}{d_i} = \frac{1}{.3333} = 3$$

or

$$\tilde{\sigma}_a^2 = \frac{6.175}{3} = 2.058$$

and

$$V\left(\tilde{\sigma}_a^2\right) = \frac{2k_1}{k_0^2}\sigma_a^4 + \frac{4}{k_0^2}\sigma_a^2\sigma^2 + \frac{2}{k_0^2}\left[4 + \frac{16}{9}\right]\sigma^4$$

where $k_1 = \sum \frac{1}{d_i^2} = 9$ or

$$V\left(\tilde{\sigma}_a^2\right) = 2\sigma_a^4 + \frac{4}{3}\sigma_a^2\sigma^2 + \frac{104}{81}\sigma^4$$

$$= 2\sigma_a^4 + 1.33\sigma_a^2\sigma^2 + 1.28\sigma^4$$

It is clear that $\hat{\sigma}_a^2$ has a smaller variance than $\tilde{\sigma}_a^2$. By the above procedure, we can generate a rather large set of unbiased estimates and find the variances of the estimates in a rather efficient manner. This approach is contrasted with Searle's paper [29] in which it took considerable effort to develop an expression just for the variance of one unbiased estimate.

6.2 The Two-Way Classification Without Interaction

6.2.1 No Missing Cells

Let us consider the example taken from Harvey [18].

		Ration No.	
		1	2
	1	5 6	2 3
Sire No.	2	2 5 7 3 6	8 9 8
	3	3	4 6 7 4 6

The model is

$$y_{ijk} = u_{ij} + e_{ijk}$$

$i = 1,\cdots,3;\ j = 1,2;\ k = 1,\cdots,n_{ij}$

subject to

$$u_{11} - u_{21} - u_{12} + u_{22} = 0$$

$$u_{11} - u_{31} - u_{12} + u_{32} = 0$$

The matrix $A(X^TX)^{-1}A^T$ is given in Section 4.2.2. Now

$$(u^*)^T = (5.5, 2.5, 4.6, 8.3, 3.0, 5.4)$$

$$(\hat{u})^T = (3.19, 4.81, 5.39, 7.01, 3.65, 5.27)$$

$$\hat{\sigma}^2 = 4.02$$

Let us test H_o: $\bar{u}_{i\cdot} = \bar{u}_{i'\cdot}$ and H_o: $\bar{u}_{\cdot j} = \bar{u}_{\cdot j'}$. The corresponding hypothesis matrices are also given in Section 4.2.2.

Now

$$SS\left(\Lambda_1^T\right) = 15.68$$

$$SS\left(\Lambda_2^T\right) = 9.707$$

and

$$F_1 = \frac{15.68}{2 \cdot (4.02)} = 1.95$$

$$F_2 = \frac{9.707}{4.02} = 2.41$$

Now suppose that before the experiment was run we thought that $u_{22} = 2u_{11}$. Therefore, another test is H_o: $u_{22} = 2u_{11}$. Most texts do not treat hypotheses like this; however, the sums of squares can be obtained by

$$SS(\Lambda^T) = (\Lambda^T\hat{u})^T[\Lambda^T A(X^TX)^{-1}A^T\Lambda]^{-1}(\Lambda^T\hat{u})$$

where $\Lambda^T = [2 \quad 0 \quad 0 \quad -1 \quad 0 \quad 0]$

or

$$SS(\Lambda^T) = \sum_{i=1}^{1} \frac{1}{d_i}\left(\lambda_i^T\hat{u}\right)^2 = \frac{(1.63)^2}{1.84} = 1.44$$

and

$$F_3 = \frac{1.44}{4.02} = .35$$

The estimates of variance components are discussed in Section 4.2.2.

6.2.2 Missing Cells

Suppose we were given the following data

Fabric		Temperature 1	Temperature 2	Temperature 3	Temperature 4
	1	No observations	1.8 2.1 2.0 2.1	4.6	7.5 7.9
	2	2.2 2.4	4.2 4.0	5.4 5.6	9.2
	3	2.8 3.2	No observations	8.7 8.4	13.2
	4	No observations	3.2 3.6 3.3	5.7 5.8	10.9 11.1

The model is

$$y_{ijk} = u_{ij} + e_{ijk}$$

subject to $\theta^T u = 0$ where

$$\theta^T =$$

$$\begin{array}{ccccccccccccc}
u_{12} & u_{13} & u_{14} & u_{21} & u_{22} & u_{23} & u_{24} & u_{31} & u_{33} & u_{34} & u_{42} & u_{43} & u_{44}
\end{array}$$

$$\begin{bmatrix}
1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & 0
\end{bmatrix}$$

Now

$$\hat{\sigma}^2 = 8.183$$

and

$$(u^*)^T = (2.0,\ 4.6,\ 7.7,\ 2.3,\ 4.1,\ 5.5,\ 9.2,\ 3.0,\ 8.55,\ 13.2,\ 3.37,\ 5.75,\ 11.0)$$

$$(\hat{u})^T = (1.84,\ 4.01,\ 8.30,\ 1.55,\ 3.82,\ 5.99,\ 10.3,\ 3.74,\ 8.17,\ 12.47,\ 3.76,\ 5.73,\ 10.22)\ .$$

Now suppose we test H_o: $u_{ij} = u_{i'j}$ and H_o: $u_{ij} = u_{ij'}$, then the appropriate sums of squares are, respectively,

$$SS\left(\Lambda_1^T\right) = 37.86$$

$$SS\left(\Lambda_2^T\right) = 215.23$$

Again we can test H_o: $\lambda^T u = 0$ for arbitrary λ^T .

6.2.3 BIB Design

The following example is taken from Graybill [15].

Total Weights of Nitrogen in Grams of Six Cuttings
of Alfalfa Forage Grown in Greenhouse Pots

Block j	Treatments			Block total $Y_{\bullet j \bullet}$
1	2.10 (1)	2.67 (2)	2.91 (4)	7.68
2	1.14 (2)	3.00 (3)	3.10 (5)	7.24
3	2.92 (3)	3.14 (4)	2.99 (6)	9.05
4	3.13 (4)	2.63 (5)	2.75 (7)	8.51
5	2.84 (5)	3.13 (6)	1.85 (1)	7.82
6	3.01 (6)	2.99 (7)	1.82 (2)	7.82
7	2.86 (7)	3.54 (1)	3.78 (3)	10.18

$k = 3,\ t = 7,\ b = 7,\ r = 3,\ \lambda = 1$

Grand total $Y_{\bullet\bullet\bullet} = 58.30$

The model is $y_{ij} = u_{ij} + e_{ij}$ subject to $\theta^T u = 0$

where

$$
\theta^T u = \begin{bmatrix}
0 & 0 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 \\
1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1
\end{bmatrix} u
$$

and where

$$
\begin{aligned}
u^T = [&u_{11}, u_{15}, u_{17}, u_{21}, u_{22}, u_{26}, u_{32}, \\
&u_{33}, u_{37}, u_{41}, u_{43}, u_{44}, u_{52}, u_{54}, \\
&u_{55}, u_{63}, u_{65}, u_{66}, u_{74}, u_{76}, u_{77}]
\end{aligned}
$$

Now $\hat{\sigma}^2 = .2424$ and

$$u^* = \begin{bmatrix} 2.0999999 \\ 1.8500000 \\ 3.5399999 \\ 2.6700000 \\ 1.1400000 \\ 1.8200000 \\ 3.0000000 \\ 2.9200000 \\ 3.7800000 \\ 2.9100000 \\ 3.1400000 \\ 3.1300000 \\ 3.0999999 \\ 2.6300000 \\ 2.8399999 \\ 2.9899999 \\ 3.1300000 \\ 3.0100000 \\ 2.7500000 \\ 2.9899999 \\ 2.8600000 \end{bmatrix} \qquad \hat{u} = \begin{bmatrix} 2.4233332 \\ 2.0704761 \\ 2.9961904 \\ 2.0461904 \\ 1.6290476 \\ 1.9547619 \\ 2.8404762 \\ 3.0290476 \\ 3.8304762 \\ 3.2104761 \\ 2.9819047 \\ 2.9876190 \\ 2.7704761 \\ 2.9647619 \\ 2.8347618 \\ 3.0390476 \\ 2.9147619 \\ 3.1761904 \\ 2.5576191 \\ 2.6890476 \\ 3.3533332 \end{bmatrix}.$$

Now suppose we test

(1) H_o: $u_{ij} = u_{i'j}$

(2) H_o: $u_{ij} = u_{ij'}$

(3) $H_o: \quad \bar{u}_{i\cdot} = \bar{u}_{i'\cdot}$

(4) $H_o: \quad \bar{u}_{\cdot j} = \bar{u}_{\cdot j'}$

These are

(1) Treatment "adjusted" for blocks.

(2) Blocks "adjusted" for treatments.

(3) Treatments "unadjusted."

(4) Blocks "unadjusted."

In tabular form, we have

Source of Variation	df	SS
$H_o: \quad u_{ij} = u_{i'j}$	6	3.2708
$H_o: \quad u_{ij} = u_{ij'}$	6	1.5370
$H_o: \quad \bar{u}_{i\cdot} = \bar{u}_{i'\cdot}$	6	3.7885
$H_o: \quad \bar{u}_{\cdot j} = \bar{u}_{\cdot j'}$	6	2.0347
Error	8	1.9388

Now if we are in a fixed model and are interested in just $H_o: \quad u_{ij} = u_{i'j}$, then we see that

$$F = \frac{3.2908}{6 \cdot (2424)} = 2.263 \ .$$

And if we are in a mixed model, where treatments are fixed and blocks are random, we see that

$$E[SS(\Delta^T)] = 6\sigma^2 + 14\sigma_b^2$$

where Δ^T is associated with H_o: $u_{ij} = u_{ij'}$

Now

$$\hat{\sigma}^2 = .2424$$

$$\hat{\sigma}_b^2 = \frac{1.5375 - 6(.2424)}{14} = .006$$

Also

$$u_{ij} - u_{i'j} = \alpha_i - \alpha_{i'}$$

and

$$E(\bar{u}_{i\cdot} - \bar{u}_{i'\cdot}) = \alpha_i - \alpha_{i'}$$

Let λ_1^T be such that $\lambda_1^T u = u_{ij} - u_{i'j}$ and λ_2^T be such that $\lambda_2^T u = \bar{u}_{i\cdot} - \bar{u}_{i'\cdot}$. Now $\lambda_1^T\hat{u}$ and $\lambda_2^T\hat{u}$ both are unbiased estimates of $\alpha_i - \alpha_{i'}$. Let $\sigma_{11} = V\left(\lambda_1^T\hat{u}\right)$, $\sigma_{22} = V\left(\lambda_2^T\hat{u}\right)$ and $\sigma_{12} = \mathrm{Cov}\left(\lambda_1^T\hat{u}, \lambda_2^T\hat{u}\right)$. Then

$$V\begin{bmatrix}\lambda_1^T\hat{u} \\ \lambda_2^T\hat{u}\end{bmatrix} = \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{bmatrix} = V$$

Now from the discussion in Section 4.2.3, we know a better estimate of $\alpha_i - \alpha_{i'}$ would be a $a_1\left(\lambda_1^T\hat{u}\right) + a_2\left(\lambda_2^T\hat{u}\right)$ where

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \frac{V^{-1}\ell}{\ell^T V^{-1}\ell}$$

and where $\ell^T = [1 \quad 1]$. But σ_{11}, σ_{12} and σ_{22} are functions of σ^2 and σ_b^2 . However, if we use the $\hat{\sigma}^2$ and $\hat{\sigma}_b^2$ as described above, then $\lambda_1^T\hat{u}$ and $\lambda_2^T\hat{u}$ will be uncorrelated with $\hat{V}^{-1}$ since Δ^T satisfies the conditions described in Section 4.2.3. Hence an unbiased estimate of $\alpha_i - \alpha_{i'}$ is $a_1\left(\lambda_1^T\hat{u}\right) + a_2\left(\lambda_2^T\hat{u}\right)$ where

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \frac{\hat{V}^{-1}\ell}{\ell^T \hat{V}^{-1}\ell} \quad .$$

For example, suppose we wanted to estimate $\alpha_1 - \alpha_2$. We could use $\lambda_1^T\hat{u} = \hat{u}_{11} - \hat{u}_{21} = .377$ and

$$\begin{aligned} \lambda_2^T\hat{u} &= \tfrac{1}{2}\,[\hat{u}_{15} + \hat{u}_{17} - \hat{u}_{22} - \hat{u}_{26} + \hat{u}_{37} - \hat{u}_{32} \\ &\quad + \hat{u}_{55} - \hat{u}_{52} + \hat{u}_{65} - \hat{u}_{66} + \hat{u}_{77} - \hat{u}_{76}] \\ &= 1.47 \end{aligned}$$

Now $\hat{u} = u + Ae$ where $u = \gamma 1 + U_1\alpha + V_1 b$ and $E[\hat{u}] = \gamma 1 + U_1\alpha$.

$$V[\hat{u}] = \sigma_b^2 V_1 V_1^T + \sigma^2 A$$

Now $\lambda_1^T\hat{u}$ and $\lambda_2^T\hat{u}$ are unbiased estimates of $\alpha_1 - \alpha_2$. And

$$V\left(\lambda_1^T\hat{u}\right) = \sigma_b^2\lambda_1^T V_1 V_1^T \lambda_1 + d_{11}\sigma^2$$

where

$$d_{11} = .86$$

$$\lambda_1^T V_1 V_1^T \lambda_1 = 0$$

Hence

$$V\left(\lambda_2^T\hat{u}\right) = .86\sigma^2$$

Also

$$V\left(\lambda_2^T\hat{u}\right) = \sigma_b^2\lambda_2^T V_1 V_1^T \lambda_1 + d_{22}\sigma^2$$

where

$$d_{22} = 3$$

$$\lambda_2^T V_1 V_1 \lambda_2 = 9$$

Thus

$$V\left(\lambda_2^T\hat{u}\right) = 3\sigma^2 + 9\sigma_b^2$$

$$\begin{aligned}\text{Cov}\left(\lambda_1^T\hat{u},\ \lambda_2^T\hat{u}\right) &= \sigma_b^2\lambda_1^T V_1 V_1^T\lambda_2 + \sigma^2\lambda_1^T A(X^TX)^{-1}A^T\lambda_2 \\ &= 0\ .\end{aligned}$$

Now substituting $\hat{\sigma}^2$ and $\hat{\sigma}_b^2$ which we found above, we have

$$\hat{V}\begin{bmatrix}\lambda_1^T\hat{u}\\ \lambda_2^T\hat{u}\end{bmatrix} = \begin{bmatrix}.86\sigma^2 & 0\\ 0 & 3\sigma^2 + 9\sigma_b^2\end{bmatrix}$$

$$= \begin{bmatrix}.2085 & 0\\ 0 & .7812\end{bmatrix}$$

and $\hat{V}$ is uncorrelated with $\lambda_1^T\hat{u}$ and $\lambda_2^T\hat{u}$. Therefore,

$$\begin{pmatrix}a_1\\ a_2\end{pmatrix} = \frac{\begin{bmatrix}\frac{1}{.2085} & 0\\ 0 & \frac{1}{.7812}\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix}}{[1 \quad 1]\begin{bmatrix}\frac{1}{.2085} & 0\\ 0 & \frac{1}{.7812}\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix}}$$

$$= \frac{1}{6.076}\begin{bmatrix}4.796\\ 1.280\end{bmatrix} = \begin{bmatrix}.789\\ .211\end{bmatrix}$$

Hence another unbiased estimate of $\alpha_1-\alpha_2$ is

$$a_1\left(\lambda_1^T\hat{u}\right) + a_2\left(\lambda_2^T\hat{u}\right) = .297 + .310$$

$$= .607$$

This is the same estimate as the one obtained by Graybill [15], where he used the "interblock" and "intrablock" estimates. Let us exhibit another estimate obtained from the general procedure. Let

$$\lambda_1^T\hat{u} = \hat{u}_{11} - \hat{u}_{21} = .377$$

$$\lambda_2^T\hat{u} = \frac{1}{2}[\hat{u}_{15} - \hat{u}_{17} - \hat{u}_{22} - \hat{u}_{26} + \hat{u}_{37} - \hat{u}_{32}$$

$$+ \hat{u}_{55} - \hat{u}_{52} + \hat{u}_{65} - \hat{u}_{66} + \hat{u}_{77} - \hat{u}_{76}]$$

$$= 1.47$$

$$\lambda_3^T\hat{u} = \frac{1}{3}[u_{11} + u_{15} + u_{17} - u_{21} + u_{22} - u_{26}]$$

$$= .620$$

We see that $\lambda_1^T\hat{u}$, $\lambda_2^T\hat{u}$ and $\lambda_3^T\hat{u}$ are unbiased estimates of $\alpha_1-\alpha_2$. Now

$$V\left(\lambda_3^T\hat{u}\right) = \sigma_b^2\lambda_3^T V_1 V_1\lambda_3 + d_{33}\sigma^2$$

where

$$d_{33} = \frac{2}{3}$$

$$\lambda_3^T V_1 V_1^T \lambda_3 = \frac{16}{9}$$

Therefore,

$$V\left(\lambda_3^T \hat{u}\right) = \frac{16}{9}\,\sigma_b^2 + \frac{2}{3}\,\sigma^2$$

and

$$\mathrm{Cov}\left(\lambda_1^T \hat{u},\ \lambda_3^T \hat{u}\right) = \lambda_1^T A \lambda_3 \sigma^2 = \frac{2}{3}\,\sigma^2$$

$$\mathrm{Cov}\left(\lambda_2^T \hat{u},\ \lambda_3^T \hat{u}\right) = \lambda_2^T V_1 V_1^T \lambda_3 \sigma_b^2 + \lambda_2^T A \lambda_3 \sigma^2$$

$$= \frac{2}{3}\,\sigma^2 + 2\sigma_b^2$$

Now

$$\hat{V} = \begin{bmatrix} .2085 & 0 & .1616 \\ 0 & .7812 & .1618 \\ .1616 & .1618 & .1726 \end{bmatrix}$$

and $\hat{V}$ is uncorrelated with $\lambda_1^T \hat{u}$, $\lambda_2^T \hat{u}$ and $\lambda_3^T \hat{u}$.

Hence

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \frac{\hat{V}^{-1}\ell}{\ell^T\hat{V}^{-1}\ell}$$

$$= \begin{bmatrix} .6226 \\ .1661 \\ .2113 \end{bmatrix}$$

Thus, another unbiased estimate of $\alpha_1-\alpha_2$ is

$$a_1\left(\lambda_1^T\hat{u}\right) + a_2\left(\lambda_2^T\hat{u}\right) + a_3\left(\lambda_3^T\hat{u}\right) = .6099$$

Let us find an estimate of $\alpha_1-\alpha_2$ using the concepts developed in Section 4.2.4. There we saw that another unbiased estimate was $\lambda^T\bar{u}$ where

$$\bar{u} = [I - (X^T\hat{V}^{-1}X)^{-1}\theta[\theta^T(X^T\hat{V}^{-1}X)^{-1}\theta]^{-1}\theta^T]$$
$$\cdot [X^T\hat{V}^{-1}X]^{-1}X^T\hat{V}^{-1}Y$$

where $\hat{V}$ is an estimate of the covariance matrix of Y.

In our case, we have

$$\bar{u} = [I - \hat{V}\theta(\theta^T\hat{V}\theta)^{-1}\theta^T]Y$$

Now $\lambda^T\bar{u} = \bar{u}_{11} - \bar{u}_{21}$ is an unbiased estimate of $\alpha_1 - \alpha_2$ and $\lambda^T\bar{u} = .376$. In this case, $\hat{V}$ is almost diagonal since the nonzero off-diagonal terms are .006. Hence $\bar{u}$ and $\hat{u}$ are not significantly different.

6.3 The Two-Way Classification With Interaction

6.3.1 No Missing Cells

Let us consider the example given in Section 6.2.1. Since $\theta^T = 0$, we have that

$$(\hat{u})^T = (u^*)^T$$
$$= (5.5,\ 2.5,\ 4.6,\ 8.3,\ 3.0,\ 5.4) .$$

Now suppose we test

(1) H_o: $\bar{u}_{i\cdot} = \bar{u}_{i'\cdot}$

(2) H_o: $\bar{u}_{\cdot j} = \bar{u}_{\cdot j'}$

(3) H_o: $u_{ij} - u_{i'j} - u_{ij'} + u_{i'j'} = 0$

Note that (3) is equivalent to H_o: $\theta^T u = 0$ where θ^T is defined in Section 6.2.1.

The results are tabulated below

Hypothesis	df	SS
H_o: $\bar{u}_{i\cdot} = \bar{u}_{i'\cdot}$	2	21.00
H_o: $\bar{u}_{\cdot j} = \bar{u}_{\cdot j'}$	1	3.59
H_o: $\theta^T u = 0$	2	30.22
Error	12	26.07

To illustrate, let us find the expected value of $SS(\Lambda^T)$, where Λ^T is the matrix associated with (2). Now $\Lambda^T = \frac{1}{\sqrt{6}}[1 \ \ -1 \ \ 1 \ \ -1 \ \ 1 \ \ -1]$.

Suppose we assume the following structure on u ,

$$u = \gamma 1 + U_1\alpha + V_1 b + V_2\delta$$

where $\gamma 1$, and $U_1\alpha$ denote the fixed effect, $V_1 b$ and $V_2\delta$ denote the random effects, and where $b \sim N\left(0,\sigma_b^2 I\right)$ $\delta \sim N\left(0,\sigma_\delta^2 I\right)$. Let us note that some authors [28] assume that $\delta \sim N(0,V)$ where V has some nonzero off-diagonal elements. Naturally, their results will differ from ours. Now $\hat{u} = u + (X^TX)^{-1}X^Te$, hence

$$E[\hat{u}] = \gamma 1 + U_1\alpha$$

$$V[\hat{u}] = \sigma_b^2 V_1 V_1^T + \sigma_\delta^2 V_2 V_2^T + \sigma^2 (X^TX)^{-1}$$

Note that $V_2 = I$. Hence

$$V[\hat{u}] = \sigma_b^2 V_1 V_1^T + \sigma_\delta^2 I + \sigma^2 (X^T X)^{-1}$$

Now

$$E[SS(\Lambda^T)] = \frac{1}{d_1} [[\Lambda^T V(\hat{u})\Lambda] + (\Lambda^T u)^2]$$

Now

$$U_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad V_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Hence

$$\begin{aligned} \Lambda^T u &= \Lambda^T \gamma 1 + \Lambda^T U_1 \alpha \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} \Lambda^T V(\hat{u})\Lambda &= \left(\Lambda^T V_1 V_1^T \Lambda\right)\sigma_b^2 + \sigma_\delta^2 \Lambda^T \Lambda I + (\Lambda^T (X^T X)^{-1} \Lambda)\sigma^2 \\ &= 3\sigma_b^2 + \sigma_\delta^2 + d_1 \sigma^2 \end{aligned}$$

Hence

$$E[SS(\Lambda^T)] = \frac{3}{d_1} \sigma_b^2 + \frac{1}{d_1} \sigma_\delta^2 + \sigma^2$$

where

$$d_1 = \Lambda^T(X^TX)^{-1}\Lambda$$

$$= \frac{41}{90}$$

Hence

$$E[SS(\Lambda^T)] = \sigma^2 + 2.195\sigma_\delta^2 + 6.585\sigma_b^2$$

Note that there is a σ_δ^2 term in this equation. This does not agree with those authors who assume $\delta \sim N(0,V)$.

6.3.2 Missing Cells

We shall consider the example given in Section 6.2.2. Here $\theta^T = 0$. The major question is which hypotheses are to be tested. Let us test the following.

(1) H_o: $\bar{u}_{i\cdot} = \bar{u}_{i'\cdot}$

(2) H_o: $\bar{u}_{\cdot j} = \bar{u}_{\cdot j'}$

(3) H_o: $\theta^T u = 0$

where θ^T is defined in Section 6.2.2. Now

$$(\hat{u}) = (u^*)$$

$$= (2.0,\ 4.6,\ 7.7,\ 2.3,\ 4.1,\ 5.5,\ 9.2,\ 3.0,\ 8.55,\ 13.2,\ 3.37,\ 5.75,\ 11.0)$$

The results are tabulated below

Hypothesis	df	SS
H_o: $\bar{u}_{i\cdot} = \bar{u}_{i'\cdot}$	3	36.66
H_o: $\bar{u}_{\cdot j} = \bar{u}_{\cdot j'}$	3	201.95
H_o: $\theta^T u = 0$	6	7.7461
Error	13	0.4366

This concludes the examples on the two-way classification with interaction.

6.4 Latin Square

This example is taken from Snedecor [32].

		Cows 1	Cows 2	Cows 3
Period	1	A: 608	B: 885	C: 940
	2	B: 715	C: 1087	A: 776
	3	C: 844	A: 711	B: 832

The model is

$$y_{ijk} = u_{ijk} + e_{ijk}$$

$i = 1,2,3$; $j = 1,2,3$; $k = 0,1,n_{ij}$ subject to no interactions.

Hence $\theta^T u$ is

$$\theta^T u = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 & -1 & 0 & 1 \\ 1 & 0 & -1 & -1 & 1 & 0 & 0 & -1 & 1 \end{bmatrix}$$

where

$$u^T = [u_{111} \quad u_{122} \quad u_{133} \quad u_{212} \quad u_{223} \quad u_{231}$$
$$u_{313} \quad u_{321} \quad u_{332}]$$

Now we test the following hypotheses:

(1) H_o: $\bar{u}_{1\cdot\cdot} = \bar{u}_{2\cdot\cdot} = \bar{u}_{3\cdot\cdot}$

(2) H_o: $\bar{u}_{\cdot 1\cdot} = \bar{u}_{\cdot 2\cdot} = \bar{u}_{\cdot 3\cdot}$

(3) H_o: $\bar{u}_{\cdot\cdot 1} = \bar{u}_{\cdot\cdot 2} = \bar{u}_{\cdot\cdot 3}$

Let us consider (3) in detail.

$$\bar{u}_{\cdot\cdot 1} = \frac{u_{111} + u_{321} + u_{231}}{3}$$

$$\bar{u}_{\cdot\cdot 2} = \frac{u_{122} + u_{212} + u_{332}}{3}$$

$$\bar{u}_{\cdot\cdot 3} = \frac{u_{133} + u_{223} + u_{313}}{3}$$

In terms of the "β" model

$$u_{ijk} = u + \alpha_i + \gamma_j + \delta_k$$

Hence

$$\bar{u}_{\bullet\bullet 1} = u + \delta_1 + \frac{\gamma_1 + \gamma_2 + \gamma_3}{3} + \frac{\alpha_1 + \alpha_2 + \alpha_3}{3}$$

$$\bar{u}_{\bullet\bullet 2} = u + \delta_2 + \frac{\gamma_1 + \gamma_2 + \gamma_3}{3} + \frac{\alpha_1 + \alpha_2 + \alpha_3}{3}$$

$$\bar{u}_{\bullet\bullet 3} = u + \delta_3 + \frac{\gamma_1 \quad \gamma_2 \quad \gamma_3}{3} + \frac{\alpha_1 \quad \alpha_2 \quad \alpha_3}{3}$$

Hence H_o: $\bar{u}_{\bullet\bullet 1} = \bar{u}_{\bullet\bullet 2} = \bar{u}_{\bullet\bullet 3}$ is equivalent to H_o: $\delta_1 = \delta_2 = \delta_3$ in the "β" model. The results are given below.

$$(u^*)^T = (608, 885, 940, 715, 1087, 776, 844, 711, 832)$$

$$(\hat{u})^T = (711.3, 883.3, 759.7, 931.7, 886.7, 696, 868, 823)$$

Hypothesis	df	SS
H_o: $\bar{u}_{1\bullet\bullet} = \bar{u}_{2\bullet\bullet} = \bar{u}_{3\bullet\bullet}$	2	5900.
H_o: $\bar{u}_{\bullet 1\bullet} = \bar{u}_{\bullet 2\bullet} = \bar{u}_{\bullet 3\bullet}$	2	47214.
H : $\bar{u}_{\bullet\bullet 1} = \bar{u}_{\bullet\bullet 2} = \bar{u}_{\bullet\bullet 3}$	2	103436.
Error	2	4843.

Let us now consider the case where we have a missing cell. Suppose $y_{111} = 608$, is missing. The model remains the same except that

$$\theta^T u = [1 \quad -1 \quad -1 \quad 0 \quad 1 \quad 1 \quad -1 \quad 0]u$$

where

$$u^T = [u_{122}, u_{133}, u_{212}, u_{223}, u_{231}, u_{313}, u_{321}, u_{332}]$$

Again we have the problem as to what we should test. Let us test H_o: $\Lambda^T u = 0$ or

$$H_o: \left\{ \begin{array}{lcl} u_{212} - u_{231} - u_{313} + u_{332} & = & 0 \\ u_{223} - u_{122} - u_{231} + u_{133} & = & 0 \end{array} \right\}$$

This is equivalent to H_o: $\delta_1 = \delta_2 = \delta_3$ in the "β" model. The results are given in the following table.

Hypothesis	df	SS
H_o: $\Lambda^T u = 0$	2	80960.
Error	1	2773.0

6.5 Simultaneous Confidence Intervals

Let us consider the example in Section 6.1. That is $y_{ij} = u_i + e_{ij}$ where $i = 1,2,3,4,5$, $j = 1,2,\cdots,n_i$. Now suppose we wanted an $\alpha\%$ confidence interval for $\Lambda^T u$. From Section 2.4, we see that the set of all ξ such that

$$\frac{(\xi - \Lambda^T\hat{u})^T(\Lambda^T(X^TX)^{-1}\Lambda)^{-1}(\xi - \Lambda^T\hat{u})}{s\hat{\sigma}^2} \leq F_{(\alpha,s,n-p)}$$

is an a% confidence interval, where $F_{(\alpha,s,n-p)}$ is the $\alpha\%$ point of a central F distribution with s and $n-p$ degrees of freedom. Let us take $\alpha = .95$ and

$$\Lambda^T u = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} u = \begin{bmatrix} u_1 \\ u_4 \end{bmatrix}.$$

Now the set of ξ such that

$$\frac{(\xi - \Lambda^T\hat{u})^T(\Lambda^T(X^TX)^{-1}\Lambda)^{-1}(\xi - \Lambda^T\hat{u})}{2(.80)} \leq 4.26$$

is a 95% confidence interval for $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$.

Or we have

$$\frac{\left[\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}-\begin{pmatrix}8.0\\ 9.1\end{pmatrix}\right]^T\begin{bmatrix}5 & 0\\ 0 & \frac{1}{3}\end{bmatrix}^{-1}\left[\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}-\begin{pmatrix}8.0\\ 9.1\end{pmatrix}\right]}{1.60} \leq 4.26$$

or

$$\frac{(\xi_1 - 8.0)^2}{1.36} + \frac{(\xi_2 - 9.1)^2}{2.27} \leq 1 \ .$$

CHAPTER 7. CONCLUSION

7.0 Summary

In the preceding chapters we have developed a comprehensive theory of linear models. We have shown that by formulating the model as $y_{ijk} = u_{ij} + e_{ijk}$, where $y_{ijk} \sim N(u_{ij}, \sigma^2)$ and where certain relations may be known about the u_{ij}, we could analyze this model in a very efficient manner. It was shown that the one theory can be applied to the balanced case, the unbalanced case, the missing cell case, the case of interactions and the case of no interactions. We also showed that the one theory can be applied to both classification models and design models. By using this approach, we are able to analyze mixed models, that is we can estimate components of variance and provide estimates of fixed effects based upon information recovered from the random effects. We also applied this approach to regression and covariance models.

There are a number of advantages to using the "u" model approach. First, let us contrast it with the usual approach used to teach the balanced case. Here the student is given a set of rules and a set of equations. He is told that with these he can analyze certain designs. For example, in the two way classification with interaction, he is told that the following table is the way to analyze such a model:

Source of Variation	df	SS	MS
A	a-1	A_{yy}	$\overline{A}_{yy}$
B	b-1	B_{yy}	$\overline{B}_{yy}$
AB	(a-1)(b-1)	AB_{yy}	$\overline{AB}_{yy}$
Error	ab(r-1)	E_{yy}	$\overline{E}_{yy}$

where

$$A_{yy} = bn \sum_{i=1}^{a} (\bar{y}_{i..} - \bar{y}_{...})^2$$

$$B_{yy} = an \sum_{j=1}^{b} (\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$AB_{yy} = n \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$E_{yy} = AB_{yy} - A_{yy} - B_{yy}$$

Some intuitive justification is given as to why the degrees of freedom and sums of squares arise as they do. Other tests of hypotheses are almost never considered. Also, some authors include certain restrictions as a part of their models. For example

$$y_{ijk} = u + \alpha_i + \beta_j + \delta_{ij} + e_{ijk}$$

where

$$\sum_i \alpha_i = \sum_j \beta_i = \sum_i \delta_{ij} = \sum_j \delta_{ij} = 0$$

The table above answers only three questions; that is, it tests only three hypotheses. As we know, there are infinite hypotheses one can test. But the student is told to test the same hypotheses each time he conducts such an experiment.

This approach is especially harmful to those students from other disciplines that have had only two or three "methods" courses. In those courses, we taught them only the rules of statistics rather than the rules and concepts. As a result, they do not understand the true role of statistics or of the statistician. They think statistics is limited only to certain models that can be analyzed and only certain questions that can be answered. Using standard computer programs perpetuates this misconception. Persons laboring under this misconception are the ones who give us boxes of data cards that we as statisticians are supposed to analyze by our "mystic" methods. More importantly, these students, from their experiences in the "methods" courses, have the mistaken idea that there are only four or five classical models that are valid to statistical analysis. They might make wrong assumptions about the data generated by their experiments in an effort to make those experiments fit one of the classical models. Or they might disregard some of the data in order that their experiments will fit the balanced design of the classical statistical analysis. The result could be that a model which does not describe the experiment is used. Thus, the "u" model provides a method of analysis that enables the experimenter to pose an essentially unlimited number of hypotheses. It also allows a great amount of flexibility in choosing the model.

Second, the statistician is like a merchant – he is a merchant of models. When a scientist comes to a statistician, the scientist has conjectured that a phenomenon he has observed or will observe may fit a certain model. He asks the statistician to help formulate the model so that quantitative analysis can be performed. The statistician looks over his "stock" of models and tries to find one that approximates the model presented by the scientist. If the statistician can offer only balanced models, or if the scientist can only ask predetermined questions, their efforts degenerate into a game of data shuffling which has little or no scientific value.

Statisticians should be experts in models. They should formulate abstract models, analyze them, and investigate their advantages and limitations so that when they are called upon to help the scientist, their "stock" is such that they can be of help. They will also be able to determine whether the questions posed can be answered by a given model. Thus the second advantage is that the "u" model offers a general abstract model that can be applied to a large number of practical problems.

Finally, there are several approaches to analyzing linear models. One is the generalized inverse. Another is the "u" model approach outlined in this paper. The claim here is not that the "u" model is the only approach, rather that it is one method that is conceptually simple. It is based upon the most fundamental statistical assumptions. The student can be shown how to pose the questions, how to test them, and how to interpret them without having to understand the mathematics involved. Yet the principles they learn will apply to all types of linear models and not just to balanced models.

7.1 Future Research

We will now discuss several research topics which arose from studying the "u" model.

7.1.1 Other Models

Presently, the "stock" of models is primarily linear models. We should seek to include nonlinear models. Also, the tests of hypotheses are restricted to the form H_o: $\Lambda^T u = \xi$. We should generalize this to include hypotheses of the form H_o: $\Lambda^T u \geq \xi$, H_o: $\Lambda^T \leq \xi$, etc.

7.1.2 Variance Components

Perhaps the use of the "u" model can reveal something about the properties of the estimates of the components of variances in the mixed models.

7.1.3 Estimates of Fixed Effects in the Mixed Model

We have seen that we need not restrict ourselves to the BIB design or to the classical interblock and intrablock estimates in estimating fixed effects. However, the criteria we used in finding estimates was that the estimates must be unbiased. We should consider other criteria, such as minimum variance and so on.

BIBLIOGRAPHY

[1] Anderson, R. L. and Crump, P. P., "Comparisons of Designs and Estimation Procedures for Estimating Parameters in a Two-Stage Nested Process," Technometrics, 9 (1967).

[2] Bartlett, M. S., "The Use of Transformations," Biometrics, 3 (1947).

[3] Bross, Irwin, "Fiducial Intervals for Variance Components," Biometrics, 6 (1950).

[4] Brownlee, K. A., "Statistical Theory and Methodology," John Wiley and Sons, New York, (1960).

[5] Bush, N. and Anderson, R. L., "A Comparison of Three Different Procedures for Estimating Variance Components," Technometrics, 5 (1963).

[6] Cochran, W. G., "Some Consequences When the Assumptions for the Analysis of Variance Are Not Satisfied," Biometrics, 5 (1947).

[7] Cramer, H., "Mathematical Methods of Statistics," Princeton University Press (1946).

[8] Crump, S. L., "The Estimation of Variance Components in Analysis of Variance," Biometrics, 2 (1946).

[9] Crump, S. L., "The Present Status of Variance Component Analysis," Biometrics, 7 (1951).

[10] Daniels, H. E., "The Estimation of Components of Variance," Supplement to the Journal of the Royal Statistical Society, 6 (1939).

[11] Eisenhart, C., "The Assumptions Underlying Analysis of Variance," Biometrics, 3 (1947).

[12] Elston, R. C. and Bush, N., "The Hypotheses That Can Be Tested When There Are Interactions in an Analysis of Variance Model," Biometrics, 20 (1964).

[13] Fisher, R. A., "Statistical Methods for Research Workers," 13th edition, Oliver and Boyd, Ltd., London and Edinburgh.

[14] Goulden, C. H., "Methods of Statistical Analysis," John Wiley and Sons, New York, (1939).

[15] Graybill, F. A., "An Introduction to Linear Statistical Models," McGraw-Hill Book Co., Inc., New York, (1961).

[16] Green, J. R., "A Confidence Interval for Variance Components, "Annals of Mathematical Statistics, 25 (1954).

[17] Hartley, H. O., "Expectations, Variances and Covariances of ANOVA Mean Squares by 'Synthesis'," Biometrika, 23 (1967).

[18] Harvey, W. R., "Least-Squares Analysis of Data with Unequal Subclass Numbers," U.S.D.A., ARS-20-8, (1960).

[19] Henderson, C. R., "Estimation of Variance and Covariance Components," Biometrics, 9 (1953).

[20] Jacobsen, R. L., "Linear Algebra and ANOVA," Unpublished Ph. D. Thesis, Cornell University, (1968).

[21] Koch, G. G., "A General Approach to the Estimation of Variance Components," Technometrics, 9 (1967).

[22] Kruskal, N., "The Coordinate-Free Approach to Gauss-Markov Estimation, and its Application to Missing and Extra Observations," Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1 (1960).

[23] Lehman, E. L., "Testing Statistical Hypotheses," John Wiley, New York, (1959).

[24] Ostle, B., "Statistics in Research," Iowa State University Press, 2nd edition, (1963).

[25] Rao, J. N. K., "On Expectations, Variances and Covariances of ANOVA Mean Squares by 'Synthesis'," to appear in Biometrics.

[26] Rao, C. R., "Advanced Statistical Methods in Biometric Research," John Wiley, New York, (1952).

[27] Rao, C. R., "Linear Statistical Inference and Its Application," John Wiley and Sons, New York, (1965).

[28] Scheffe, H., "The Analysis of Variance," John Wiley and Sons, New York, (1959).

[29] Searle, S. R., "Matrix Methods in Components of Variance and Covariance Analysis," Annals of Mathematical Statistics, 27 (1956).

[30] Searle, S. R., Private Communication

[31] Searle, S. R., "Matrix Algebra for the Biological Sciences," John Wiley and Sons, New York, (1966).

[32] Snedecor, G. W., "Statistical Methods: Applied to Experiments in Agriculture and Biology," The Collegiate Press, Inc., Ames, Iowa, (1934).

[33] Satterthwaithe, F. E., "An Approximate Distribution of Estimates of Variance Components," Biometrics, 2 (1946).

[34] Tippett, L. H. C., "The Methods of Statistics," Williams and Norgate, Ltd., London, (1937).

[35] Tukey, J. W., "Variances of Variance Components: II, The Unbalanced Single Classification," Annals of Mathematical Statistics, 28 (1957).

[36] Wilks, S. S., "Mathematical Statistics," John Wiley and Sons, New York, (1962).

VITA

The author was born in San Antonio, Texas, on March 1, 1942. He completed high school at Corpus Christi Academy. In 1964, he graduated cum laude from St. Mary's University of San Antonio, Texas with a B.S. in Mathematics. In 1965, he received an M.S. in Mathematics from St. Mary's. From 1964 to 1966, Mr. Speed was employed by the National Aeronautics and Space Administration, Manned Spacecraft Center, Houston, Texas. In 1966, he was granted leave to pursue a Ph.D. in Statistics at Texas A&M University. In June 1968, Mr. Speed returned to NASA. He now is in charge of the Statistics Section of the Theory and Analysis Office in the Computation and Analysis Division, MSC. In October, he was awarded the W. S. Connor Award by the Institute of Statistics, Texas A&M University. The author's parents are Mr. and Mrs. E. W. Speed, Sr. of Corpus Christi, Texas. The author resides at 1910 Bonanza, Houston, Texas 77058. The typists for this dissertation were:

1) Mrs. Kay Burns
2) Mrs. Diana Matthews