COMPONENTS OF VARIANCE ANALYSIS

COMPONENTS OF VARIANCE ANALYSIS

By

RONALD E. WALPOLE, B.A.

A Thesis

Submitted to the Faculty of Arts and Science

in Partial Fulfilment of the Requirements

for the Degree

Master of Arts

McMaster University

October 1955

MASTER OF ARTS (1955)
(Mathematics)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Components of Variance Analysis

AUTHOR: Ronald E. Walpole, B.A. (McMaster University)

SUPERVISOR: Professor J.D. Bankier

NUMBER OF PAGES: 159

SCOPE AND CONTENTS:

In this thesis a systematic and short method for computing the expected values of mean squares has been developed. One chapter is devoted to the theory of regression analysis by the method of least squares using matrix notation and a proof is given that the method of least squares leads to an absolute minimum, a result which the author has not found in the literature. For two-way classifications the results have been developed for proportional frequencies, a subject which again has been neglected in the literature except for the Type II model. Finally, the methods for computing the expected values of the mean squares are applied to nested classifications and Latin square designs.

CONTENTS

CHAPTER I

INTRODUCTION

CHAPTER II

REGRESSION ANALYSIS

CHAPTER III

ONE-WAY CLASSIFICATION MODELS

CHAPTER IV

TWO-WAY CROSSED CLASSIFICATION MODELS

CHAPTER V

TWO-WAY NESTED CLASSIFICATION MODELS

CHAPTER VI

CHAPTER I

INTRODUCTION

If a set of observations can be classified according to one or more criteria, then the total variation between the members of the set can be broken up into components which can be attributed to the different criteria of classification. By testing the significance of these components it is then possible to determine which of the criteria are associated with a significant proportion of the overall variation. To carry out the analysis it is necessary to assume for the data a model which involves a number of parameters and properties.

1.1 The One-Way Classification

Our data might be the results obtained from dye trials on each of 5 preparations of naphthalene black 12B made from each of 6 samples of H acid intermediate as recorded in Table 1.1.

TABLE 1.1[1]

Yields of Napthalene Black 12B

Sample of H acid	1	2	3	4	5	6
Individual yields in grams of standard colour.	1440	1490	1510	1440	1515	1445
	1440	1495	1550	1445	1595	1450
	1520	1540	1560	1465	1625	1455
	1545	1555	1595	1545	1630	1480
	1580	1560	1605	1595	1635	1520

[1]Bennett and Franklin, Statistical Analysis in Chemistry and the Chemical Industry. New York: John Wiley & Sons, 1954, p.320.

The data is classified according to the sample of acid used. We denote an observation by Y_{ij} $(i=1,2,\ldots,6; j=1,2,\ldots,5)$ where i indicates the number of the acid sample and j the number of the observation for this sample. These observations are considered to be random variables with the expected value

$$E(Y_{ij}) = \mu + \tau_i \quad ,$$

where τ_i is the contribution of the ith acid sample and μ is a constant. If we assume that the τ_i's are constant, we have what is known as the <u>Type I model</u> and any conclusions we might draw would apply only to our six acid samples. If we wished the conclusions to hold for a larger group of acid samples of which our six acid samples were a sample, we would consider the τ_i's themselves to be random variables. If we assume that they are drawn from an infinite population, $f(\tau_i)$ with variance σ_τ^2, we have the <u>Type II model</u>. If the population is finite, we say we have a <u>Type III model</u>. Thus the nature of the conclusions we wish to draw determines the form of the mathematical model used. It is customary to write

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where ε_{ij}, a random variable, denotes the difference between Y_{ij} and its mean value.

<u>1.2 Two-Way Crossed Classifications</u>

It is often desirable to collect experimental data so that they may be classified according to two factors. In this case our model would be

(1.21) $$Y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk}$$

when τ_i is the contribution to the mean of the ith level of the first classification and β_j is the contribution of the jth level of the second classification. For example, five workmen might take turns working on four machines. Then Y_{ijk} would be the number of articles produced on machine i by workman j on the kth day.

In setting up the above model we have assumed that the τ_i's and β_j's, the effects due to machines and workmen, were additive. If we had reason to doubt this, we would use the model

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \tag{1.22}$$

where $(\tau\beta)_{ij}$ is the interaction term associated with the ith machine and jth workman.

If all the parameters involved in model (1.21) are considered to be constants, we have a Type I model and our conclusions would apply to only the workmen and the machines used in the experiment. In this case, we would be interested in estimating the parameters and testing hypotheses about these parameters. If we wish our conclusions to apply to a larger population of machines and workmen, we would consider all the parameters, with the exception of μ, to be random variables. We have a Type II model if the random variables are assumed to come from continuous populations and a Type III model if they are assumed to come from finite populations. In addition to these three models, we may set up mixed models where some of the parameters are constants and others are random variables. The model used depends on the objectives of the experimenter.

1.3 Two-Way Nested Classification

In the two-way crossed classification it was assumed that each level of a given classification made a definite contribution to the mean of Y_{ijk}. This is not a realistic assumption for certain types of experiments. For example, in section 1.1, we consider the yields of naphthalene black for six different samples of H acid. Suppose these six samples were random samples of H acid produced from naphthalene supplied by a particular tar distiller. Suppose further that the experiment was carried out four times, the supplier of naphthalene being changed for each experiment. If we attempt to describe the data by model (1.21) or (1.22), we might regard τ_i $(i = 1,2,3,4)$ as the contribution made by the different suppliers, but it would not be reasonable to regard β_j $(j = 1,2,\ldots,6)$ as the contributions of the six random samples of acid since this would suggest that all samples having the same number have the same effect upon the yield. What is required is a model of the form

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk}$$

where τ_i is the effect upon the yield of the ith supplier and $\beta_{j(i)}$ represents the effect due to variations within samples from the ith supplier. Again, assumptions made about the parameters can lead to a variety of different models.

1.4 Additional Assumptions

It will be shown in the chapters that follow that we can impose certain linear conditions upon the parameters without loss of generality. These conditions simplify the mathematical analysis. Also it will be assumed, when the parameters are considered to be random variables, that parameters represented by different Greek letters are independently distributed. Their variances, σ_τ^2, σ_β^2, $\sigma_{\tau\beta}^2$, and σ^2 for the ε_{ijk}, are called components of variance since they are parts of the variance of the Y_{ijk}. One of our principal problems is the estimation of the components of variance. Finally, we shall always assume that the ε_{ijk} have a normal distribution and, depending on the model under consideration, distributions will be assumed for any other random parameters.

1.5 The Scope of This Thesis

In the next chapter the theory of regression analysis is presented in a form which will cover all the Type I models considered in the thesis and which will provide a foundation for the other types of models. In the third chapter the one-way classification is considered for the case of unequal numbers of observations in the different levels of the classification. The case where the number of observations are equal is then obtained as a special case and Type I, II, and III models are considered. In the remaining chapters models involving more than one classification are considered. Time and space do not permit the consideration of all possible models but it is hoped that the methods used here will be of use in developing the theory of models which have not been considered.

CHAPTER II

REGRESSION ANALYSIS

2.1 The Model

In this chapter we shall consider the problem of estimating the value of some random variable Y with a mean depending on certain variables $X_1, X_2, \ldots, X_r$, whose values may be determined exactly when Y is observed. If $n > r$ observations are made, we obtain the sets of values $(X_{1\alpha}, X_{2\alpha}, \ldots, X_{r\alpha}; Y_\alpha)$, $(\alpha = 1, 2, \ldots, n)$. If $X_1, X_2, \ldots, X_r$ were held fixed at the values $X_{1\alpha}, X_{2\alpha}, \ldots, X_{r\alpha}$, the observed value of Y, Y_α, would vary in a random fashion about its mean value which we assume has the form

$$(2.11) \qquad E(Y_\alpha) = \mu' + \sum_{i=1}^{r} \beta_i X_{i\alpha} \quad .$$

It is convenient to introduce the variables

$$x_{i\alpha} = X_{i\alpha} - \bar{X}_i \qquad (i = 1, 2, \cdots, r) ,$$

where

$$\bar{X}_i = \frac{1}{n} \sum_{\alpha=1}^{n} X_{i\alpha} \quad .$$

Then equation (2.11) may be written in the form

$$(2.12) \qquad E(Y_\alpha) = \mu + \sum_{i=1}^{r} \beta_i x_{i\alpha}$$

where

$$\mu = \mu' + \sum_{i=1}^{r} \beta_i \bar{X}_i$$

This is equivalent to saying that

$$(2.13) \qquad Y_\alpha = \mu + \sum_{i=1}^{r} \beta_i x_{i\alpha} + \varepsilon_\alpha ,$$

where ε_α is defined by this equation and is called the <u>true error</u>. A consequence of (2.12) is that $E(\varepsilon) = 0$.

Our objective is to estimate μ and the β_i $(i = 1, 2, \ldots, r)$ by the method of least squares and we will denote these estimates by $\hat{\mu}$ and b_i $(i = 1, 2, \ldots, r)$, where the b_i's are called the <u>regression coefficients</u>. We can then write

$$(2.14) \qquad Y_\alpha = \hat{\mu} + \sum_{i=1}^{r} b_i x_{i\alpha} + e_\alpha \qquad (\alpha = 1, 2, \cdots, n)$$

where e_α is called the <u>residual</u>.

The sets of equations (2.13) and (2.14) may be written in the form of the matrix equations

$$(2.15) \qquad Y = \mu + X\beta + \varepsilon = \hat{\mu} + Xb + e$$

where

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu \\ \mu \\ \cdot \\ \cdot \\ \mu \end{pmatrix} = \mu \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix} = \mu I ,$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_r \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_r \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{pmatrix}, \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{pmatrix}$$

and $X = (x_1, x_2, \ldots, x_r)$ is an $n \times r$ matrix where x_i is the column vector with elements $x_{i\alpha}$ ($\alpha = 1,2,\ldots,n$). We shall assume that the rank of X is r. Since

$$I'x_i = \sum_{\alpha=1}^{n} (X_{i\alpha} - \bar{X}_i) = 0 \quad , \quad I'X = 0 \quad .$$

Let $Z = (z_1, z_2, \ldots, z_r)$ denote an $n \times r$ matrix. Then there exists a matrix

$$W = \begin{pmatrix} w_{11} & w_{21} & w_{31} & \cdots & w_{r1} \\ 0 & w_{22} & w_{32} & \cdots & w_{r2} \\ 0 & 0 & w_{33} & \cdots & w_{r3} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & w_{rr} \end{pmatrix}$$

such that $Z = XW$, where the r column vectors of Z are orthogonal and of unit length.[1] Hence Z is of rank r. Since the rank of the product of two matrices does not exceed the rank of either of the matrices, the rank of W is also r, and W^{-1} exists. Since $I'X = 0$, $I'Z = I'X\,W = 0W = 0$. Also $Z'Z = I_r$, the $r \times r$ identity matrix, and $A = X'X = (W')^{-1}Z'Z\,W^{-1} = (WW')^{-1}$. Thus, $WW' = A^{-1}$, and we also note that $A' = A$. We have $Z = XW$ and $Z' = W'X'$ so that $I_r = Z'Z = W'X\,X'W = W'A\,W$.
Therefore $W^{-1} = W'A$.

The matrix equations (2.15) may now be written in the form

$$Y = \mu + ZW^{-1}\beta + \varepsilon = \hat{\mu} + ZW^{-1}b + e$$

---

[1]Schreier and Sperner, Modern Algebra and Matrix Theory. New York: Chelsea Publishing Company, 1951, p. 141.

or

(2.16) $$Y = \mu + Z\gamma + \varepsilon = \hat{\mu} + Zc + e$$

where

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_r \end{pmatrix} = W^{-1}\beta \quad , \quad c = \begin{pmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_r \end{pmatrix} = W^{-1}b \; .$$

2.2 Estimation of the Regression Coefficients

The least squares estimates of the scalars are obtained by minimizing the error sum of squares, SSE. From (2.15) the error sum of squares is

$$SSE = \sum_{\alpha=1}^{n} e_\alpha^2 = e'e = (Y-\hat{\mu}-Xb)'(Y-\hat{\mu}-Xb)$$

$$= (Y-\hat{\mu})'(Y-\hat{\mu}) - 2(Y-\hat{\mu})'Xb + b'X'Xb$$

(2.21) $$= (Y-\hat{\mu})'(Y-\hat{\mu}) - 2(Y-\hat{\mu})'Xb + b'Ab$$

since a 1×1 matrix is equal to its own transpose. Setting the partial derivative with respect to $\hat{\mu}$ equal to zero we obtain the equation,

$$\frac{\partial SSE}{\partial \hat{\mu}} = -2I'(Y-\hat{\mu}) - 2I'Xb = -2n(\bar{Y}-\hat{\mu}) = 0 \; ,$$

since $I'X = 0$, and therefore

(2.22) $$\hat{\mu} = \bar{Y}$$

Taking the partial derivatives with respect to b_i ($i = 1, 2, \ldots, r$) we obtain the equations,

$$\frac{\partial SSE}{\partial b_i} = -2(Y-\bar{Y})'X\frac{\partial b}{\partial b_i} + \frac{\partial b'}{\partial b_i}Ab + b'A\frac{\partial b}{\partial b_i} = 0$$

where $\frac{\partial b}{\partial b_i}$ is equal to an $r \times 1$ column matrix with a one occurring in the ith row and zeros in all the other rows. Since all the matrices for $i = 1, 2, \ldots, r$ are one element matrices, and $A' = A$, we have

$$\frac{\partial b'}{\partial b_i}Ab = b'A\frac{\partial b}{\partial b_i}$$

and

(2.23) $$\left(\frac{\partial b}{\partial b_i}\right)'Ab = \left(\frac{\partial b}{\partial b_i}\right)'X'y \, ,$$

where $y = Y - \bar{Y}$. Since Ab and $X'y$ are $r \times 1$ matrices and $\left(\frac{\partial b}{\partial b_i}\right)'$ is the $1 \times r$ row matrix where a one occurs in the ith column and zeros in all the other columns, the <u>normal</u> <u>equations</u> (2.23) may be written in the matrix form

$$Ab = X'y$$

and hence

(2.24) $$b = A^{-1}X'y,$$

where A^{-1} exists since X' is of rank r and the Gram matrix $A = X'X$ must have the same rank.

If our model is considered in the form

$$Y = \mu + Z\gamma + \varepsilon = \hat{\mu} + Zc + e \, ,$$

we must replace X by Z in the above results and we would still obtain $\hat{\mu} = \bar{Y}$ since

$$I'Z = I'XW = OW = 0,$$

which is a $1 \times r$ zero matrix. To obtain the normal equations, we would replace $A = X'X$ by $Z'Z = I_r$ obtaining

$$c = Z'y = W'X'y = W'A \ A \ X'y = W'A \ b \ .$$

Since $W'A = W^{-1}, c = W^{-1}b$. Hence, we could have obtained c by substituting for b in $\gamma = W^{-1}\beta$.

(2.3) Properties of the Regression Coefficients

We will now show that the c_i's give us a unique absolute minimum. Replacing X by Z and $\hat{\mu}$ by $\bar{Y}$ in (2.21), the error sum of squares is

(2.31) $$SSE = y'y - 2y'Zc + c'c \ .$$

Also

$$(c - Z'y)'(c - Z'y) = (c' - y'Z)(c - Z'y)$$
$$= c'c - 2y'Z \ c + y'Z \ Z'y \ .$$

Therefore

(2.32) $$SSE = (c - Z'y)'(c - Z'y) + y'y - y'ZZ'y \ .$$

This expression has a unique absolute minimum for $c = Z'y$, which is $y'y - y'ZZ'y$, since $(c - Z'y)'(c - Z'y)$ is the length of the vector $c - Z'y$.

To express this absolute minimum in terms of the original model, we make use of the relations $Z = XW$, $c = W'Ab$, $WW' = A^{-1}$, and substitute in (2.32) to obtain

$$\begin{aligned} SSE &= (b'AW - y'XW)(W'Ab - W'X'y) + y'y - y'XWW'X'y \\ &= (b'A - y'X)WW'(Ab - X'y) + y'y - y'XA^{-1}X'y \\ &= (b' - y'XA^{-1})AA \ A(b - A^{-1}X'y) + y'y - y'XA^{-1}X'y \\ &= (b - A^{-1}X'y)'A(b - A^{-1}X'y) + y'y - y'XA^{-1}X'y \\ &= u'Au + y'y - y'XA^{-1}X'y \end{aligned}$$

where $u = (b - A^{-1}X'y)$.

Now $u'Au$ is a positive definite[1] quadratic form and has a unique minimum, 0 , when

$$u = b - A^{-1}X'y = 0 .$$

This shows that SSE has a unique minimum of $y'y - y'XA^{-1}X'y$ when $b = A^{-1}X'y$.

We have

$$b = A^{-1}X'y = A^{-1}X'(X\beta + \varepsilon) = \beta + A^{-1}X'\varepsilon .$$

Thus

$$E(b) = \beta + A^{-1}X'E(\varepsilon) = \beta$$

and b_i is an unbiased estimate of β_i $(i = 1, 2, \ldots, r)$. Replacing X by Z in the above argument shows that $E(c) = \gamma$. Before computing the variance-covariance matrix of the b_i's, we introduce the additional assumption that the ε_α's are independently distributed and have a common variance, σ^2, that is $E(\varepsilon\varepsilon') = \sigma^2 I_n$. The variance-covariance matrix of the b_i's is

$$E[(b-\beta)(b-\beta)'] = E[A^{-1}X'\varepsilon\varepsilon'XA^{-1}]$$
$$= \sigma^2 A^{-1}X'I_n XA^{-1} = \sigma^2 A^{-1}AA^{-1} = \sigma^2 A^{-1} .$$

Thus the b_i's will, in general, be correlated. On the other hand the c_i's are uncorrelated, since, replacing X by Z, we find their variance-covariance matrix is $\sigma^2 I_r$, and each c_i has variance σ^2.

---

[1] Attridge, R.F., Linear Regression and Multiple Classification Designs. Hamilton: unpublished thesis, 1952, p. 63 .

(2.4) Reduction due to Regression

If no attempt were made to estimate the mean of Y,

$$\sum_{\alpha=1}^{n} Y_\alpha^2$$

would be called the sum of squares for Y and it could be thought of as giving a measure of the spread of the observations about the value $Y=0$. If it is only assumed that $E(Y)=\mu$, the least squares estimate of μ is $\bar{Y}$, and

$$\sum_{\alpha=1}^{n} (Y_\alpha - \bar{Y})^2 = \sum_{\alpha=1}^{n} y_\alpha^2$$

is a measure of the spread of the observations about $\bar{Y}$, our estimate of E(Y) . Finally, with our model (2.12) , SSE is our measure of this spread. The difference between SSE and the sum of squares, when we assume $E(Y)=\mu$, is denoted by

$$SSR = \sum_{\alpha=1}^{n} y_\alpha^2 - SSE$$

and is called the reduction in the sum of squares attributable to the regression. If SSR differs by little from $\sum_{\alpha=1}^{n} y_\alpha^2$, SSE is small, and the Y_α's are close to the estimates of their means, indicating that little is to be gained by introducing additional terms into the regression equation (2.12) .

(2.5) Tests of Hypotheses

We will be interested in testing the hypothesis that $\beta_i = 0 \ (i=1,2,\ldots,r)$. This is equivalent to the hypothesis that μ is a satisfactory expression for E(Y) . The above remarks suggest that the ratio SSR to SSE would be small when this is the case and large when the hypothesis should be rejected. In practice we use the more convenient statistic

$$F = \frac{MSR}{MSE} ,$$

where

$$MSR = \frac{SSR}{r} , \qquad MSE = \frac{SSE}{n - r - 1} .$$

The question then arises as to how large F should be if we are to reject the hypothesis. To answer this question, we next determine the distribution of F.

We now assume that the ε_α's are normally and independently distributed with mean 0 and variance σ^2, indicated by saying that the ε_α's are NID$(0,\sigma^2)$. From (2.6),

$$e = y - Zc = y - Z\gamma - Z(c - \gamma) ,$$

and

(2.51) $SSE = e'e$

$$= (y - Z\gamma)^2 - 2(y - Z\gamma)'Z(c - \gamma) + (c - \gamma)'Z'Z(c - \gamma)$$

$$= (y - Z\gamma)^2 - 2(y'Z - \gamma'Z'Z)(c - \gamma) + (c - \gamma)'(c - \gamma)$$

$$= (y - Z\gamma)^2 - 2(c' - \gamma')(c - \gamma) + (c - \gamma)'(c - \gamma)$$

$$= (y - Z\gamma)^2 - (c - \gamma)^2 .$$

Since the scalar $\bar{Y} = \frac{I'Y}{n} = \mu + \frac{1}{n} I'Z\gamma + \bar{\varepsilon} = \mu + \bar{\varepsilon}$,

$$y = Y - \bar{Y} = \mu + Z\gamma + \varepsilon - \mu - \bar{\varepsilon} = Z\gamma + \varepsilon - \bar{\varepsilon} ,$$

where

$$\bar{Y} = \begin{pmatrix} \bar{Y} \\ \bar{Y} \\ \cdot \\ \cdot \\ \bar{Y} \end{pmatrix} , \qquad \bar{\varepsilon} = \begin{pmatrix} \bar{\varepsilon} \\ \bar{\varepsilon} \\ \cdot \\ \cdot \\ \bar{\varepsilon} \end{pmatrix} .$$

Therefore $y - Z\gamma = \varepsilon - \bar{\varepsilon}$ and the error sum of squares becomes

$$(2.52) \qquad SSE = (\varepsilon - \bar{\varepsilon})^2 - (c - \gamma)^2$$
$$= \sum_{\alpha=1}^{n} (\varepsilon_\alpha - \bar{\varepsilon})^2 - \sum_{i=1}^{r} (c_i - \gamma_i)^2 .$$

We shall make frequent use of the following results. If we have a set of variables $x_1, x_2, \ldots, x_n$, which are NID(μ, σ^2),

$$\chi^2 = \sum_{\alpha=1}^{n} \left(\frac{x_\alpha - \mu}{\sigma}\right)^2$$

has a χ^2 distribution with n degrees of freedom. We often say that $\sum_{\alpha=1}^{n} (x_\alpha - \mu)^2$ is distributed as $\chi^2\sigma^2$ with n degrees of freedom (d.f.) . If μ is replaced by $\bar{x}$, the mean of the x_α's, the resulting expression has a $\chi^2\sigma^2$ distribution with n - 1 d.f. . We also use the relations

$$E(\chi^2\sigma^2) = \nu\sigma^2 \quad , \quad Var(\chi^2\sigma^2) = 2\nu\sigma^4$$

where the second expression is the variance of $\chi^2\sigma^2$ and ν is the number of d.f. associated with χ^2. Finally a statistic

$$F = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} \quad ,$$

where χ_1^2 and χ_2^2 have χ^2 distributions with ν_1 and ν_2 degrees of freedom and are independently distributed, has what is known as the F distribution with ν_1 and ν_2 degrees of freedom.

Since $c - \gamma = Z'\varepsilon$, which implies that the c_i's are independently distributed, the c_i's are NID $(0, \sigma^2)$. Since the ε_α's and the c_i's both have this property, both sums on the right of (2.52) are distributed as $\chi^2\sigma^2$, the first with n - 1 d.f. and the second with r d.f. . Hence

$$E(SSE) = (n-1)\sigma^2 - r\sigma^2 = (n-r-1)\sigma^2$$

and

$$E\left(\frac{SSE}{n - r - 1}\right) = \sigma^2 .$$

Thus

$$s^2 = \frac{SSE}{n - r - 1}$$

is an unbiased estimate of σ^2. We have seen, following (2.32), that

$$SSE = y'y - y'ZZ'y = y'y - c'c ,$$

making use of the normal equations $c = Z'y$. Thus

$$SSR = c'c = \sum_{i=1}^{r} c_i^2 .$$

We wish to test the hypothesis H: $\gamma_1 = \gamma_2 = \cdot \cdot \cdot = \gamma_r = 0$. Since $\sum_{i=1}^{r} (c_i - \gamma_i)^2$ is distributed as $\chi^2\sigma^2$, SSR is also distributed as $\chi^2\sigma^2$ with r d.f. under the null hypothesis. If the hypothesis is not true SSR/σ^2 has a non-central χ^2 distribution[1].

Now we will prove that SSE is distributed as $\chi^2\sigma^2$ with $n - r - 1$ degrees of freedom and is statistically independent of SSR. Augment the r orthogonal vectors $z_1, z_2, \ldots, z_r$ by $n - r - 1$ others, which we shall designate by $p_1, p_2, \ldots, p_{n-r-1}$, such that $z_1, \ldots, z_r, p_1, \ldots, p_{n-r-1}$ form an orthogonal set of unit vectors. The estimation equations will now be

---

[1]Patnaik, P.B., The Non-Central χ^2- and F - Distributions and their Applications, Biometrika, 36 (1949), p. 202.

(2.53) $$Y = \mu + Z\gamma + \varepsilon = \bar{Y} + Zc + Pd + f ,$$

where

$$P = (p_1, p_2, \dots, p_{n-r-1}) ,$$

$$d = \begin{pmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_{n-r-1} \end{pmatrix} \quad \text{and} \quad f = \begin{pmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_n \end{pmatrix} .$$

We may write (2.53) in the form

(2.54) $$y = Zc + Pd + f = (Z,P)\begin{pmatrix} c \\ d \end{pmatrix} + f .$$

Since (Z,P) is an orthogonal matrix, our earlier theory shows that

$$\begin{pmatrix} c \\ d \end{pmatrix} = (Z,P)'y = \begin{pmatrix} Z' \\ P' \end{pmatrix} y = \begin{pmatrix} Z'y \\ P'y \end{pmatrix} .$$

Therefore, $c = Z'y$ and $d = P'y$. We saw earlier that z_i was orthogonal to the $n \times 1$ matrix I and this is also true for p_i. From (2.54)

$$c = Z'y = Z'Zc + Z'Pd + Z'f ,$$

or $c = c + Z'f$ and $Z'f = 0$. Similarly $P'f = 0$ and $If = 0$. These n equations in n unknowns have an $n \times n$ orthogonal matrix and hence $f = 0$ or

$$y = Z\ c + P\ d .$$

Hence, $(y - Zc - Pd)^2 = 0$, that is,

$$\begin{aligned} 0 &= [y - Z\gamma - Z(c - \gamma) - Pd]^2 \\ &= (y - Z\gamma)^2 - (c - \gamma)^2 + d'P'Pd - 2y'Pd \\ &= (y - Z\gamma)^2 - (c - \gamma)^2 + d'd - 2d'd , \end{aligned}$$

making use of (2.51). Thus

$$d'd = (y - Z\gamma)^2 - (c - \gamma)^2 = SSE .$$

Hence we have broken SSE into n-r-1 orthogonal squares. Furthermore,

$$d = P'y = P'Y = P'(\mu + Z\gamma + \varepsilon) = P'\varepsilon ,$$

$E(d) = 0$, and the variance-covariance matrix of d is
$E(dd') = E(P'\varepsilon\,\varepsilon'P) = \sigma^2 P'I_n P = \sigma^2 P'P = \sigma^2 I_{n-r-1}$.
Therefore $\sigma^2(d_j) = \sigma^2$, $\sigma(d_j d_k) = 0$, $j \neq k$.
The covariances for d_j, c_i are given by

$$E[d(c-\gamma)'] = E(P'\varepsilon\varepsilon'Z) = \sigma^2 P'Z = 0 .$$

Therefore $\sigma^2(d_j c_i) = 0$

Since the d_j's are orthogonal linear forms in the ε_α's which are NID$(0,\sigma^2)$, they are NID$(0,\sigma^2)$. Hence the d_j^2 are independently distributed as $\chi^2\sigma^2$ with 1 degree of freedom each, or SSE is distributed as $\chi^2\sigma^2$ with n-r-1 degrees of freedom. Also the d_j's and $(c_i - \gamma_i)$'s, being uncorrelated and normally distributed, are independently distributed. Therefore

$$SSE = \sum_{j=1}^{n-r-1} d_j^2 \qquad \text{and} \qquad SSR = \sum_{i=1}^{r} c_i^2$$

are independently distributed. Under these conditions, the statistic

$$F = \frac{MSR}{MSE} = \frac{\sum_{i=1}^{r} c_i^2}{r} \Big/ \frac{SSE}{n-r-1}$$

has what is known as the F distribution with r and n-r-1 d.f. . To test the hypothesis, we select a number α, $0 < \alpha < 1$, called the <u>level of significance</u>, and determine from the tables the value $F_{1-\alpha}$ such that the probability of F exceeding $F_{1-\alpha}$ is α. When this occurs we reject the hypothesis. Thus when the hypothesis is true, we would reject it on the average 100α percent of the time. We should also notice that

$$t = \frac{(c_i - \gamma_i)/\sigma}{\sqrt{s^2/\sigma^2}} = \frac{c_i - \gamma_i}{s}$$

has a t distribution since $(c_i - \gamma_i)/\sigma$ is N(0,1) and s^2/σ^2 is distributed as χ^2 with n-r-1 degrees of freedom. This statistic can be used to test the hypothesis that γ_i has a specified value.

Now suppose we assume $\gamma_1, \gamma_2, \cdots, \gamma_k \neq 0$ and test the null hypothesis that $\{\gamma_{k+1}, \gamma_{k+2}, \cdots, \gamma_r = 0\}$. Let SSR_k be the reduction in the sum of squares due to $\gamma_1, \gamma_2, \cdots, \gamma_k$ and SSR_r that due to $\gamma_1, \gamma_2, \cdots, \gamma_r$. In the first case

$$SSE = \sum_{\alpha=1}^{n} y_\alpha^2 - \sum_{i=1}^{k} c_i^2$$

and in the second case

$$SSE = \sum_{\alpha=1}^{n} y_\alpha^2 - \sum_{i=1}^{r} c_i^2 .$$

Thus

$$SSR_k = \sum_{i=1}^{k} c_i^2 , \qquad SSR_r = \sum_{i=1}^{r} c_i^2 .$$

The additional reduction in the sum of squares due to the introduction of $\gamma_{k+1}, \gamma_{k+2}, \cdots, \gamma_r$ is

$$SSR_r - SSR_k = \sum_{i=k+1}^{r} c_i^2 .$$

If this reduction is large we would reject our hypothesis that $\{\gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_r = 0\}$. We have shown that the $(c_i - \gamma_i)^2$ we independently distributed as $\chi^2\sigma^2$ with 1 degree of freedom each. Hence, under the null hypothesis that $\{\gamma_{k+1}, \gamma_{k+2}, \cdots, \gamma_r = 0\}$,

$$SSR_r - SSR_k = \sum_{i=k+1}^{r} c_i^2$$

is distributed as $\chi^2\sigma^2$ with r-k degrees of freedom. We have seen that

$$SSE = \sum_{j=1}^{n-r-1} d_j^2$$

is distributed as $\chi^2\sigma^2$ with n-r-1 d.f. and that the c_i's and d_j's are independently distributed. Hence SSE and $SSR_r - SSR_k$ are independently distributed and

$$F = \frac{SSR_r - SSR_k}{(r-k)\sigma^2} \Big/ \frac{SSE}{(n-r-1)\sigma^2} \tag{2.55}$$

has the F distribution with r-k and n-r-1 d.f. . A knowledge of σ^2 is not required since it cancels out in the computation of F. This is the statistic we use to test the hypothesis that $\{\gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_r = 0\}$

In practice, the hypothesis to be tested will be that $\{\beta_{k+1}, \beta_{k+2}, \ldots, \beta_r = 0\}$. We now show that this is equivalent to the hypothesis that $\{\gamma_{k+1}, \gamma_{k+2}, \cdots, \gamma_r = 0\}$. We have $\beta = W\gamma$ where

$$W = \begin{pmatrix} w_{11} & w_{21} & w_{31} & \cdots & w_{r1} \\ 0 & w_{22} & w_{32} & \cdots & w_{r2} \\ 0 & 0 & w_{33} & \cdots & w_{r3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & w_{rr} \end{pmatrix},$$

and $|W| = w_{11}w_{22}\cdots w_{rr} \neq 0$. Then

$$\beta_j = \sum_{i=j}^{r} w_{ij}\gamma_i \qquad (j = 1, 2, \cdots, r) ,$$

and $\{\gamma_{k+1}, \gamma_{k+2}, \cdots, \gamma_r = 0\}$ implies that $\{\beta_{k+1}, \beta_{k+2}, \cdots, \beta_r = 0\}$. Since $w_{ii} \neq 0$ $(i = 1,2,\ldots,r)$, the above equations can be solved for the γ_i's in terms of the β_i's and we obtain equations of the same form. Hence $\{\beta_{k+1}, \beta_{k+2}, \cdots, \beta_r = 0\}$ implies that $\{\gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_r = 0\}$. It follows that we may still use the statistic F of (2.17) to test our hypothesis.

While the matrix Z exists, it is difficult to construct it and hence to obtain the c_i's and the statistic F. Accordingly we need to obtain SSR_k and SSR_r in terms of the original observations. The expression SSE is the unique absolute minimum of $\sum_{\alpha=1}^{n} e_\alpha^2$ and we saw that it was obtained whether we worked in terms of the original or the orthogonal model. Hence

$$SSR_r = \sum_{\alpha=1}^{n} y_\alpha^2 - SSE$$

does not depend on the model. Now we consider

$$SSR_k = \sum_{\alpha=1}^{n} y_\alpha^2 - SSE'$$

where SSE' is the minimum of $\sum_{\alpha=1}^{n} e_\alpha^2$ when $\{\gamma_{k+1}, \gamma_{k+2}, \cdots, \gamma_r = 0\}$ But we have seen that this implies that $\{\beta_{k+1}, \beta_{k+2}, \cdots, \beta_r = 0\}$ so the above conclusions still hold and we will obtain the same value for SSR_k with either model.

Finally, we derive a result which will be needed in later chapters. We saw in section 2.3 that $c - \gamma = Z'\varepsilon$

Hence the $(c_i - \gamma_i)$'s are linear combinations of NID variables and thus are normally distributed. Also their variance-covariance matrix is $\sigma^2 I_r$. Hence the joint distribution of the c_i's is the multivariate normal distribution

$$f(c_1, c_2, \cdots, c_r) = \frac{1}{(\sqrt{2\pi}\,\sigma)^r}\, e^{-\frac{1}{2\sigma^2}(c-\gamma)'(c-\gamma)}$$

We also found that the c_i's were distributed independently of SSE so that

$$f(c_1,c_2,\ldots,c_r,SSE) = f(c_1,c_2,\ldots,c_r)f(SSE) .$$

Since $c = W^{-1}b$,

$$\begin{aligned} f(b_1,b_2,\ldots,b_r,SSE) &= f(c_1,c_2,\ldots,c_r,SSE)\,|W^{-1}| \\ &= f(c_1,c_2,\ldots,c_r)f(SSE)\,|W^{-1}| \\ &= f(b_1,b_2,\ldots,b_r)f(SSE) . \end{aligned}$$

Hence the b_i's are distributed independently of SSE. It also follows that the b_i's have a multivariate normal distribution with the β_i's as means and $\sigma^2 A^{-1}$ as the variance-covariance matrix.

CHAPTER III

ONE-WAY CLASSIFICATION MODELS

3.1 The Type I Model for Unequal Numbers per Cell

Suppose we consider a planned experiment in which p different treatments are applied to N different experimental units such that the first treatment is applied to n_1 of these units, the second to n_2 of these units and, in general, the ith treatment is applied to n_i of the units $(i = 1,2,\ldots,p)$. That is, we have divided the N experimental units into p portions of size $n_1, n_2, \ldots, n_p$. These units are usually called plots and the numbers, n_i, are often called the number of replications for the corresponding treatments. It is our purpose to test the yield-producing ability of the different treatments. The yield for the ith treatment on the jth plot of the n_i plots associated with this particular treatment, could be estimated by the model

$$(3.11) \qquad Y_{ij} = \mu' + \tau_i' + \varepsilon_{ij} ,$$

where $(j = 1,2,\ldots,n_i)$. The parameter τ_i' is the differential effect of the ith treatment over the mean μ'. We wish to test the null hypothesis that $\{\tau_1' = \tau_2' = \cdots, = \tau_p'\}$. Equation (3.11) can also be written in the form

$$(3.12) \qquad Y_{ij} = \mu' + \sum_{k=1}^{p} \tau_k' X_{ki} + \varepsilon_{ij} ,$$

where

$$X_{ki} = 0 \quad \text{for} \quad k \neq i ,$$
$$= 1 \quad \text{for} \quad k = i .$$

If we denote the mean of the τ_i''s by

$$\bar{\tau}' = \sum_{i,j} \frac{\tau_i'}{N} = \frac{1}{N} \sum_{i=1}^{p} n_i \tau_i' ,$$

and set $\tau_k = \tau_k' - \bar{\tau}'$ in equation (3.12), we obtain

$$Y_{ij} = \mu + \sum_{k=1}^{p} \tau_k X_{ki} + \varepsilon_{ij} , \tag{3.13}$$

where

$$\mu = \mu' + \bar{\tau}' \sum_{k=1}^{p} X_{ki} = \mu' + \bar{\tau}'$$

We now have the restriction that

$$\sum_{k=1}^{p} n_k \tau_k = \sum_{k=1}^{p} n_k \tau_k' - \sum_{k=1}^{p} n_k \bar{\tau}' = N\bar{\tau}' - N\bar{\tau}' = 0 .$$

If we order the Y_{ij}'s in some way, calling them Y_α ($\alpha = 1,2, \ldots, N$), the equations (3.13) can be written in the form

$$Y = \mu + \sum_{k=1}^{p} \tau_k X_k + \varepsilon . \tag{3.14}$$

Also,

$$X_i \cdot X_k = \delta_{ik} n_i , \quad \sum_{k=1}^{p} X_k = I \quad \text{and} \quad X_i \cdot I = n_i .$$

We define

$$\bar{X}_k = \frac{\sum_{\alpha=1}^{N} X_{k\alpha}}{N} = \sum_{i,j} \frac{X_{ki}}{N} = \sum_{i=1}^{p} \frac{n_i X_{ki}}{N} = \frac{n_k}{N} ,$$

and $x_{k\alpha} = X_{k\alpha} - \bar{X}_k$, so that

$$\sum_{\alpha=1}^{N} x_{k\alpha} = \sum_{i=1}^{p} n_i x_{ki} = 0 .$$

We also denote by $\bar{X}_k$ the vector $\bar{X}_k I$ and set $x_k = X_k - \bar{X}_k$. Substituting in (3.14), we obtain

$$Y = \mu + \sum_{k=1}^{p} \tau_k (x_k + \bar{X}_k) + \varepsilon \tag{3.15}$$

$$= \mu + \sum_{k=1}^{p} \tau_k \bar{X}_k + \sum_{k=1}^{p} \tau_k x_k + \varepsilon$$

$$= \mu + \sum_{k=1}^{p} \tau_k x_k + \varepsilon \quad ,$$

since
$$\sum_{k=1}^{p} n_k \tau_k = 0 \quad .$$

We can also write (3.15) in the form

(3.16) $$Y = \mu + X\tau + \varepsilon \quad ,$$

where

$$X = (x_1, x_2, \ldots, x_p)$$

and

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_p \end{pmatrix} \quad .$$

We have

$$\sum_{k=1}^{p} x_k = \sum_{k=1}^{p} X_k - \sum_{k=1}^{p} \bar{X}_k = 0$$

since

$$\sum_{k=1}^{p} \bar{X}_k = I \sum_{k=1}^{p} \frac{n_k}{N} = I \quad .$$

Thus the x_k's are not linearly independent and the rank of the matrix X is not p as required in Chapter II. To meet this difficulty, we use the relation

$$\sum_{k=1}^{p} n_k \tau_k = 0$$

to eliminate τ_p from (3.15), obtaining

$$Y = \mu + \sum_{k=1}^{p-1} \left(x_k - \frac{n_k}{n_p} x_p \right) \tau_k + \varepsilon \quad . \tag{3.17}$$

Note that

$$x_k - \frac{n_k}{n_p} x_p = X_k - \bar{X}_k - \frac{n_k}{n_p}(X_p - \bar{X}_p)$$

$$= X_k - \frac{n_k}{n_p} X_p - \frac{n_k}{N} I + \frac{n_k n_p}{n_p N} I$$

$$= X_k - \frac{n_k}{n_p} X_p \ .$$

Consider the equation

$$\sum_{k=1}^{p-1} c_k \left(X_k - \frac{n_k}{n_p} X_p \right) = 0 \quad .$$

Multiplying the equation by X_i $(i = 1, 2, \ldots, p-1)$, we find that $n_i c_i = 0$, since the X_k's are orthogonal vectors. Thus the vectors

$$x_k - \frac{n_k}{n_p} x_p \qquad (k = 1, 2, \ldots, p-1)$$

are linearly independent and our model is of the form given in Chapter II.

As before, it is assumed that the ε_{ij}'s are NID$(0, \sigma^2)$. We can now use the theory of Chapter II to test the hypothesis that $\{\tau_1 = \tau_2 = \cdots = \tau_{p-1} = 0\}$, which is equivalent to testing our original hypothesis that $\{\tau_1 = \tau_2 = \cdots = \tau_p = 0\}$ since

$$\sum_{k=1}^{p} n_k \tau_k = 0 .$$

The statistic used to make the test is

$$F = \frac{MST}{MSE} = \frac{SST/(p-1)}{SSE/(N-p)} ,$$

where SSE is obtained by minimizing the residual sum of squares

$$\sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - m - t_i)^2 ,$$

and

$$SST = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - SSE$$

is the reduction in the sum of squares attributable to the regression. In order to compute SST and SSE it is necessary to estimate the values of the parameters involved.

3.2 Estimation of the Parameters

We wish to estimate the values of μ and the τ_i's from the experiment and to do so it is necessary to select n_i plots for each treatment at random. Each treatment will then have the same chance of appearing on a given plot. Also, the randomization allows us to assume the errors to be uncorrelated. We could obtain our least squares estimates of $\mu, \tau_1, \tau_2, \cdots, \tau_{p-1}$ from the equation (3.17)

$$Y = \mu + \sum_{k=1}^{p-1} \left(x_k - \frac{n_k}{n_p} x_p \right) \tau_k + \varepsilon ,$$

but a more convenient and equivalent method is to us the method of Lagrange multipliers on the equation (3.14)

$$Y = \mu + \sum_{k=1}^{p} X_k \tau_k + \varepsilon$$

with the side condition

$$(3.21) \qquad \sum_{k=1}^{p} n_k \tau_k = 0 \quad .$$

This means that we minimize the expression

$$(3.22) \qquad \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - m - t_i)^2 + \lambda \sum_{k=1}^{p} n_k t_k \quad ,$$

where $m, t_1, t_2, \ldots, t_p$ are our least squares estimates of μ, $\tau_1, \tau_2, \ldots, \tau_p$. If we make use of the following theorem[1], we can omit the second term in (3.22).

Theorem 3.1:

If $E(Y) = \mu I + \sum_{k=1}^{p} X_k \tau_k$

and (1) $I = \sum_{k=1}^{s} X_k \quad , \quad s \leq p \quad ,$

(2) $X_1, X_2, \ldots, X_s$ form a mutually orthogonal set,

(3) $\sum_{k=1}^{s} n_k \tau_k = 0 \;, \quad \sum_{k=1}^{s} n_k \neq 0 \quad ,$

(4) any number of other conditions hold for $\tau_{s+1}, \tau_{s+2}, \ldots, \tau_p$, such that the method of Lagrange multipliers may be used, then condition (3) may be ignored. That is, its Lagrange multiplier is equal to zero in the minimizing of

---

[1] Mann, H.B., Analysis and Design of Experiments. New York: Dover Publications, 1949, p. 39.

$$SSE = (Y - \mu I - \sum_{\kappa=1}^{p} X_\kappa \tau_\kappa)^2 .$$

To determine m and the t_i's we set the partial derivatives of

$$(3.23) \qquad SSE = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - m - t_i)^2 ,$$

with respect to m and t_i, equal to zero, obtaining the normal equations

$$\frac{\partial SSE}{\partial m} = \sum_{i,j} Y_{ij} - Nm - \sum_{i=1}^{p} n_i t_i$$

$$= \sum_{i,j} Y_{ij} - Nm = 0 ,$$

$$\frac{\partial SSE}{\partial t_i} = \sum_{j=1}^{n_i} Y_{ij} - n_i m - n_i t_i = 0 \qquad (i = 1, 2, \cdots, p) .$$

From now on we shall replace a subscript by a dot to indicate summation over that subscript and we shall represent the corresponding mean by the addition of a bar. In terms of this notation, our normal equations yield the solution

$$m = \frac{1}{N} \sum_{i,j} Y_{ij} = \frac{Y_{..}}{N} = \bar{Y}_{..} ,$$

and

$$t_i = \bar{Y}_{i.} - m = \bar{Y}_{i.} - \bar{Y}_{..} .$$

3.3 Reduction due to Regression

Substituting our estimates of the parameters μ and τ_i in (3.23), the error sum of squares becomes

$$SSE = \sum_{i,j} (Y_{ij} - \bar{Y}_{i.})^2 = \sum_{i,j} Y_{ij}^2 - \sum_{i=1}^{p} n_i \bar{Y}_{i.}^2$$

$$= \sum_{i,j} y_{ij}^2 - \left(\sum_{i=1}^{p} n_i \bar{Y}_{i.}^2 - N \bar{Y}_{..}^2 \right)$$

$$= \sum_{i,j} y_{ij}^2 - \sum_{i=1}^{p} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 ,$$

where

$$y_{ij} = Y_{ij} - \bar{Y}_{..} \quad .$$

Hence the reduction in the sum of squares due to treatments is

$$SST = \sum_{i=1}^{p} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 .$$

In terms of the original model,

$$\bar{Y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mu + \tau_i + \varepsilon_{ij}) = \mu + \tau_i + \bar{\varepsilon}_{i.} , \tag{3.31}$$

and

$$\bar{Y}_{..} = \frac{1}{N} \sum_{i,j} (\mu + \tau_i + \varepsilon_{ij}) = \mu + \frac{1}{N} \sum_{i,j} \tau_i + \bar{\varepsilon}_{..}$$

$$= \mu + \frac{1}{N} \sum_{i=1}^{p} n_i \tau_i + \bar{\varepsilon}_{..} = \mu + \bar{\varepsilon}_{..} \quad . \tag{3.32}$$

Therefore

$$SST = \sum_{i=1}^{p} n_i (\tau_i + \bar{\varepsilon}_{i.} - \bar{\varepsilon}_{..})^2 .$$

In what follows, many of the expressions for the sum of squares will be of this form and it will be convenient to introduce a theorem which permits us to write down the expected value immediately. This theorem is a generalization of one due to Tukey.[1]

Theorem 3.2:

If $y_1, y_2, \ldots, y_p$ have means $\mu_1, \mu_2, \cdots, \mu_p$, variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2$ and every pair has the same covariance, λ, then

$$E\left\{ \sum_{i=1}^{p} n_i (y_i - \bar{y}.)^2 \right\} = \sum_{i=1}^{p} n_i (\mu_i - \bar{\mu}.)^2 + \sum_{i=1}^{p} n_i \left(1 - \frac{n_i}{N}\right)\left(\sigma_i^2 - \lambda\right)$$

where

$$\bar{y}. = \frac{1}{N} \sum_{i=1}^{p} n_i y_i \quad , \quad \bar{\mu}. = \frac{1}{N} \sum_{i=1}^{p} n_i \mu_i \quad ,$$

and

$$N = \sum_{i=1}^{p} n_i \quad .$$

Proof: We have

$$\sigma_i^2 = E[(y_i - \mu_i)^2] = E(y_i^2) - \mu_i^2$$

and

$$\lambda = E[(y_i - \mu_i)(y_{i'} - \mu_{i'})] = E(y_i y_{i'}) - \mu_i \mu_{i'}$$

so that

$$E(y_i^2) = \mu_i^2 + \sigma_i^2 \,, \quad E(y_i y_{i'}) = \mu_i \mu_{i'} + \lambda \quad , \quad i \neq i' .$$

Also

$$\sum_{i=1}^{p} n_i (y_i - \bar{y}.)^2 = \sum_{i=1}^{p} n_i y_i^2 - N \bar{y}.^2 \quad .$$

---

[1] Tukey, J.W., Dyadic Analysis of Variance, Human Biology, 21 (1949), pp. 65-110 .

Then

$$E(\bar{y}.) = \frac{1}{N}\sum_{i=1}^{p} n_i E(y_i) = \bar{\mu}. \quad ,$$

and

$$\begin{aligned} E(\bar{y}.^2) &= \frac{1}{N^2}\sum_{i,i'} n_i n_{i'} E(y_i y_{i'}) \\ &= \frac{1}{N^2}\sum_{i=1}^{p} n_i^2 E(y_i^2) + \frac{1}{N^2}\sum_{i\neq i'} n_i n_{i'} E(y_i y_{i'}) \\ &= \frac{1}{N^2}\sum_{i=1}^{p} n_i^2(\mu_i^2+\sigma_i^2) + \frac{1}{N^2}\sum_{i\neq i'} n_i n_{i'}(\lambda+\mu_i\mu_{i'}) \\ &= \frac{1}{N^2}\sum_{i,i'} n_i n_{i'}\mu_i\mu_{i'} + \frac{1}{N^2}\sum_{i=1}^{p} n_i^2\sigma_i^2 + \frac{\lambda}{N^2}\sum_{i\neq i'} n_i n_{i'} \ . \end{aligned}$$

But

$$N^2 = \sum_{i,i'} n_i n_{i'} = \sum_{i=1}^{p} n_i^2 + \sum_{i\neq i'} n_i n_{i'} \quad ,$$

so that

$$E(N\bar{y}.^2) = N\bar{\mu}.^2 + \frac{1}{N}\sum_{i=1}^{p} n_i^2\sigma_i^2 + \frac{\lambda}{N}\left(N^2 - \sum_{i=1}^{p} n_i^2\right) .$$

Hence,

$$\begin{aligned} E\left\{\sum_{i=1}^{p} n_i(y_i-\bar{y}.)^2\right\} &= \sum_{i=1}^{p} n_i(\mu_i^2+\sigma_i^2) - N\bar{\mu}.^2 - \frac{1}{N}\sum_{i=1}^{p} n_i^2\sigma_i^2 \\ &\quad -\lambda\left(N - \frac{1}{N}\sum_{i=1}^{p} n_i^2\right) \\ &= \sum_{i=1}^{p} n_i(\mu_i-\bar{\mu}.)^2 + \sum_{i=1}^{p} n_i\left(1-\frac{n_i}{N}\right)(\sigma_i^2-\lambda) \ . \end{aligned}$$

<u>Corollary 3.21</u>: Setting $n_i = n$ and hence $N = np$ in theorem 3.2 we obtain

$$E\{\sum_{i=1}^{p}(y_i-\bar{y}.)^2\} = (p-1)(\bar{\sigma}.^2-\lambda)+\sum_{i=1}^{p}(\mu_i-\bar{\mu}.)^2 ,$$

where

$$\bar{y}. = \frac{1}{p}\sum_{i=1}^{p} y_i \quad , \quad \bar{\mu}. = \frac{1}{p}\sum_{i=1}^{p}\mu_i$$

and

$$\bar{\sigma}.^2 = \frac{1}{p}\sum_{i=1}^{p}\sigma_i^2 \quad .$$

We shall now use theorem 3.2 to find the <u>expected treatment sum of squares</u>, E(SST). From equation (3.31) we have

$$E(\bar{Y}_i.) = \mu + \tau_i \quad ,$$

$$Var(\bar{Y}_i.) = E(\bar{\varepsilon}_i.^2) = \frac{\sigma^2}{n_i} = \sigma_i^2 \, ,$$

and

$$cov(\bar{Y}_i.,\bar{Y}_{i'}.) = E(\bar{\varepsilon}_i.\,\bar{\varepsilon}_{i'}.) = 0 = \lambda \quad .$$

Also

$$\bar{\mu}. = \frac{1}{N}\sum_{i=1}^{p} n_i(\mu+\tau_i) = \mu + \frac{1}{N}\sum_{i=1}^{p} n_i\tau_i = \mu \, .$$

Therefore the expected treatment sum of squares is

$$E(SST) = E\{\sum_{i=1}^{p} n_i(\bar{Y}_i. - \bar{Y}..)^2\}$$

$$= \sum_{i=1}^{p} n_i\tau_i^2 + \sum_{i=1}^{p} n_i(1-\frac{n_i}{N})\frac{\sigma^2}{n_i}$$

$$= \sum_{i=1}^{p} n_i\tau_i^2 + (p-1)\sigma^2 \, .$$

Hence, the expected value of the mean sum of squares due to treatments is given by

$$E(MST) = E\frac{(SST)}{p-1} = \sigma^2 + \frac{1}{p-1}\sum_{i=1}^{p} n_i \tau_i^2 .$$

We have seen that

$$SSE = \sum_{i,j}(Y_{ij} - \bar{Y}_{i.})^2$$

$$= \sum_{i,j}(\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2 .$$

Therefore the expected sum of squares due to error, by corollary 3.21, is given by

$$E(SSE) = E\left\{\sum_{i,j}(\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2\right\} = \sum_{i=1}^{p} E\left\{\sum_{j=1}^{n_i}(\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2\right\}$$

$$= \sum_{i=1}^{p}(n_i - 1)\sigma^2 = (N-p)\sigma^2$$

since

$$E(\varepsilon_{ij}) = 0 , \quad Var(\varepsilon_{ij}) = \sigma^2 , \quad and \quad cov(\varepsilon_{ij}, \varepsilon_{ij'}) = 0 .$$

Hence, the expected value of the mean sum of squares due to error is

$$E(MSE) = \frac{E(SSE)}{N-p} = \sigma^2 .$$

We could also have obtained this result by using the theory of Chapter II. Then SSE is distributed as $\chi^2\sigma^2$ with N-p degrees of freedom and hence E(SSE) = (N-p) σ^2. From Chapter II we know that SST is distributed as $\chi^2\sigma^2$ with p-1 d.f. under the null hypothesis. The analysis-of-variance table is:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Treatments	p-1	$SST=\sum_{i=1}^{p} n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2$	$MST=\frac{SST}{p-1}$	$\sigma^2+\sum_{i=1}^{p}\frac{n_i\tau_i^2}{p-1}$
Error	N-p	$SSE=\sum_{i,j}(Y_{ij}-\bar{Y}_{i.})^2$	$MSE=\frac{SSE}{N-p}$	σ^2
Total	N-1	$\sum_{i,j}(Y_{ij}-\bar{Y}_{..})^2$		

More convenient formulas for the computation of the sums of squares are

$$SST=\sum_{i=1}^{p}\frac{Y_{i.}^2}{n_i}-\frac{Y_{..}^2}{N}$$

and

$$\sum_{i,j}(Y_{ij}-\bar{Y}_{..})^2=\sum_{i,j}Y_{ij}^2-\frac{Y_{..}^2}{N},$$

while SSE is obtained by subtraction.

In the particular case where $n_i = n$ $(i = 1,2,\ldots,p)$, we define

$$\sigma_\tau^2=\sum_{i=1}^{p}\frac{\tau_i^2}{p-1}$$

If our hypothesis holds, $\sigma_\tau^2=0$. Then,

$$(SST+SSE)/\sigma^2=\sum_{i,j}(Y_{ij}-\bar{Y}_{..})^2/\sigma^2$$

has a χ^2 distribution with N-1 d.f. and

$$\sum_{i,j}(Y_{ij}-\bar{Y}_{..})^2/N-1$$

is an unbiased estimate of σ^2. If our hypothesis is rejected, we estimate σ_τ^2 and σ^2 by solving the equations

$$\sigma^2 + n\sigma_\tau^2 = MST \;, \qquad \sigma^2 = MSE \;.$$

3.4 Components of Variance for the Type II Model

Let us estimate the yield for the ith treatment on the jth plot of the n_i plots associated with this particular treatment by the model

(3.41) $$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \;,$$

where the τ_i's and ε_{ij}'s are assumed to be NID with means zero and variances σ_τ^2 and σ^2 respectively. The assumption of normality is not required for the purpose of estimating the parameters. However, this assumption is required if the usual tests of significance and confidence limits are used. From (3.41) we have

$$E(Y_{ij}) = \mu \;,$$

and

$$Var(Y_{ij}) = Var(\mu + \tau_i + \varepsilon_{ij}) = Var(\tau_i + \varepsilon_{ij})$$

$$= E(\tau_i + \varepsilon_{ij})^2 = \sigma_\tau^2 + \sigma^2 \;.$$

The type of experiment that we are concerned with here will be quite different from that in the previous sections. Here we want to estimate the mean, μ, and the variance of this estimate and to obtain estimates of the variance components, σ^2 and σ_τ^2. Our estimate of the mean is to be applicable to a wider area than that of the plots used in the experiment.

We shall arbitrarily begin with the sums of squares obtained in the previous section and show that they may be used to estimate σ_τ^2 and σ^2. In terms of the original model,

$$\bar{Y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mu + \tau_i + \varepsilon_{ij}) = \mu + \tau_i + \bar{\varepsilon}_{i.} \quad , \tag{3.42}$$

and

$$\bar{Y}_{..} = \frac{1}{N} \sum_{i,j} (\mu + \tau_i + \varepsilon_{ij}) = \mu + \frac{1}{N} \sum_{i=1}^{p} n_i \tau_i + \bar{\varepsilon}_{..} \quad . \tag{3.43}$$

From equation (3.42) we have

$$E(\bar{Y}_{i.}) = \mu \quad ,$$

$$Var(\bar{Y}_{i.}) = E[(\tau_i + \bar{\varepsilon}_{i.})^2] = \sigma_\tau^2 + \frac{\sigma^2}{n_i} = \sigma_i^2 \quad ,$$

and

$$\begin{aligned} cov(\bar{Y}_{i.}, \bar{Y}_{i'.}) &= cov(\mu + \tau_i + \bar{\varepsilon}_{i.} , \mu + \tau_{i'} + \bar{\varepsilon}_{i'.}) \\ &= E[(\tau_i + \bar{\varepsilon}_{i.})(\tau_{i'} + \bar{\varepsilon}_{i'.})] = 0 = \lambda \quad . \end{aligned}$$

Also

$$\bar{\mu}_. = \frac{1}{N} \sum_{i=1}^{p} n_i \mu = \mu$$

Therefore the expected treatment sum of squares is

$$\begin{aligned} E(SST) &= E\left\{ \sum_{i=1}^{p} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \right\} \\ &= \sum_{i=1}^{p} n_i \left(1 - \frac{n_i}{N}\right)\left(\sigma_\tau^2 + \frac{\sigma^2}{n_i}\right) \\ &= \left(\sum_{i=1}^{p} n_i - \frac{1}{N} \sum_{i=1}^{p} n_i^2 \right) \sigma_\tau^2 + \sum_{i=1}^{p} \left(1 - \frac{n_i}{N}\right) \sigma^2 \end{aligned}$$

$$= \left(N - \frac{1}{N}\sum_{i=1}^{p} n_i^2\right)\sigma_\tau^2 + (p-1)\,\sigma^2 .$$

Hence the expected value of the mean sum of squares due to treatments is given by

$$(3.44) \qquad E(MST) = \frac{E(SST)}{p-1} = \sigma^2 + \frac{1}{p-1}\left(N - \frac{1}{N}\sum_{i=1}^{p} n_i^2\right)\sigma_\tau^2 .$$

We also have

$$SSE = \sum_{i,j}(Y_{ij} - \bar{Y}_{i.})^2 = \sum_{i,j}(\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2 ,$$

and as before

$$E(SSE) = (N-p)\,\sigma^2 .$$

Therefore

$$E(s^2) = \frac{E(SSE)}{N-p} = \sigma^2 .$$

3.5 Distributions of the Sums of Squares

Corresponding to our hypothesis $\tau_i = 0$ $(i = 1,2,\ldots,p)$ for the Type I model, we have here the hypothesis $\sigma_\tau = 0$. Since $E(\tau) = 0$, this is equivalent to saying that τ will always have the value zero and that our treatments have no effect. When this hypothesis is true SST and SSE have the same values as they had in the case of the Type I model and hence are independently distributed as $\chi^2\sigma^2$ with p-1 and N-p d.f., respectively. As before, SSE has a $\chi^2\sigma^2$ distribution with N-p d.f., whether the hypothesis is true or not, since it does not depend on the τ_i's. Thus the test used for the Type I model may still be used. We could then estimate σ^2 by pooling SST and SSE and their degrees of freedom, as was done for the Type I model, when $\sigma_\tau = 0$. If the hypothesis is rejected, σ^2 and σ_τ^2 may be estimated by equating MST and MSE to their expected values.

In considering the Type I model, we found that SST/σ^2 had a non-central χ^2 distribution when the hypothesis was false. We shall now show that, under certain conditions, SST is distributed as χ^2 after it is divided by E(MST).

In Chapter II we found that $t_i = \bar{Y}_{i.} - \bar{Y}_{..}$, $(i=1,2,\ldots,p-1)$, were distributed independently of

$$SSE = \sum_{i,j} (Y_{ij} - \bar{Y}_{i.})^2$$

Therefore any function of the t_i's, which includes t_p, is independently distributed of SSE. We obtained these results by making use of the fact that

$$E(Y_{ij}) = \mu + \tau_i \quad , \quad \sum_{k=1}^{p} n_k \tau_k = 0 \quad .$$

If in particular, $E(Y_{ij}) = 0$, these results will still hold. Hence, if we replace Y_{ij} by ε_{ij} we can show that the $\bar{\varepsilon}_{i.} - \bar{\varepsilon}_{..}$ are distributed independently of

$$\sum_{i,j} (\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2 \; ,$$

where $(i=1,2,\ldots,p)$. We can write the treatment sum of squares in the form

$$SST = \sum_{i=1}^{p} n_i (\bar{\varepsilon}_{i.} - \bar{\varepsilon}_{..} + \tau_i - \bar{\tau})^2 \; ,$$

where

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^{p} n_i \tau_i \quad .$$

Since the τ_i's are independent of the ε_{ij}'s, SST is distributed independently of

$$SSE = \sum_{ij} (\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2 \quad .$$

If we let $y_i = \tau_i + \bar{\varepsilon}_{i.}$, then we have

$$E(y_i) = 0 \quad \text{and} \quad Var(y_i) = E(\tau_i + \bar{\varepsilon}_{i.})^2 = \sigma_\tau^2 + \frac{\sigma^2}{n_i} \quad .$$

The covariance of the y_i's is

$$cov(y_i, y_{i'}) = cov(\tau_i + \bar{\varepsilon}_{i.}, \tau_{i'} + \bar{\varepsilon}_{i'.}) = 0 \quad .$$

Therefore the y_i's are NID$(0, \sigma_\tau^2 + \frac{\sigma^2}{n_i})$. If we consider the case when $n_i = n$, then by Chapter II,

$$\frac{SST}{n\sigma_\tau^2 + \sigma^2} = \sum_{i=1}^{p} \frac{(y_i - \bar{y})^2}{\sigma_\tau^2 + \frac{\sigma^2}{n}}$$

has a χ^2 distribution with p-1 degrees of freedom since

$$\bar{\tau} + \bar{\varepsilon}_{..} = \frac{1}{N} \sum_{ij} (\tau_i + \varepsilon_{ij}) = \frac{1}{N} \sum_{i=1}^{p} n_i \tau_i + \frac{1}{N} \sum_{ij} \varepsilon_{ij}$$

$$= \frac{n}{np} \sum_{i=1}^{p} \tau_i + \frac{n}{np} \sum_{i=1}^{p} \bar{\varepsilon}_{i.} = \frac{1}{p} \sum_{i=1}^{p} y_i = \bar{y}$$

is the unweighted arithmetic mean of the y_i's. Using the results of Chapter II, section 2.5, we find that

$$E(MST) = \sigma^2 + n\sigma_\tau^2 \quad .$$

Substituting $n_i = n$ in the equation (3.44) we obtain

$$E(MST) = \sigma^2 + \frac{1}{p-1}\left(np - \frac{n^2 p}{np}\right)\sigma_\tau^2$$

$$= \sigma^2 + \frac{n}{p-1}(p-1)\sigma_\tau^2 = \sigma^2 + n\sigma_\tau^2 \, ,$$

which agrees with the result already obtained.

The question remains as to whether the above results might still hold when the n_i's are not all equal. We shall show that this need not be the case. First we note that if SST/c has a χ^2 distribution with k d.f.,

$$E\left(\frac{SST}{c}\right) = k$$

and

$$E(MST) = E\left(\frac{SST}{k}\right) = c \quad .$$

If $p=2$,

$$\bar{y} = \frac{n_1y_1 + n_2y_2}{N} \quad , \quad y_1-\bar{y} = \frac{n_2(y_1-y_2)}{N} \quad , \quad y_2-\bar{y} = \frac{n_1(y_2-y_1)}{N} \quad ,$$

and

$$SST = \frac{(n_1n_2^2 + n_1^2n_2)(y_1-y_2)^2}{N^2} = \frac{n_1n_2(y_1-y_2)^2}{N} \quad .$$

We have

$$E(y_1 - y_2) = 0 \, , \quad Var(y_1 - y_2) = 2\sigma_r^2 + \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

and thus

$$\frac{SST}{[2n_1n_2\sigma_r^2 + \sigma^2(n_1+n_2)]/N} = \frac{(y_1 - y_2)^2}{2\sigma_r^2 + \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

has a χ^2 distribution with one d.f. .

Thus, for $p=2$, the result still holds but we shall now show that it need not hold for $p=3$. We have

$$SST = \sum_{i=1}^{p} n_i\left[y_i - \sum_{j=1}^{p} \frac{n_j y_j}{N}\right]^2$$

$$= \sum_{i=1}^{p} n_i \left[\sum_{j=1}^{p} \left(\delta_{ij} - \frac{n_j}{N} \right) y_j \right]^2$$

$$= \sum_{j=1}^{p} \sum_{i=1}^{p} n_i \left(\delta_{ij} - \frac{n_j}{N} \right)^2 y_j^2 + \sum_{j \neq k}^{p} \sum_{i=1}^{p} n_i \left(\delta_{ij} - \frac{n_j}{N} \right)\left(\delta_{ik} - \frac{n_k}{N} \right) y_j y_k .$$

Also

$$\sum_{i=1}^{p} n_i \left(\delta_{ij} - \frac{n_j}{N} \right)^2 = n_j \left(1 - \frac{n_j}{N} \right)^2 + \sum_{\substack{i,j \\ i \neq j}} \frac{n_i n_j^2}{N^2}$$

$$= n_j \left(1 - \frac{n_j}{N} \right)^2 + \frac{n_j^2}{N^2} (N - n_j) = \frac{n_j}{N} (N - n_j) ,$$

and

$$\sum_{i=1}^{p} n_i \left(\delta_{ij} - \frac{n_j}{N} \right)\left(\delta_{ik} - \frac{n_k}{N} \right) = - \frac{n_j n_k}{N^2} (2N - n_j - n_k) + \sum_{\substack{i=1 \\ i \neq j,k}} n_i \frac{n_j n_k}{N^2}$$

$$= \frac{n_j n_k}{N^2} (2N - n_j - n_k) + \frac{n_j n_k}{N^2} (N - n_j - n_k) = - \frac{n_j n_k}{N} .$$

Thus

$$SST = \sum_{j=1}^{p} n_j \left(1 - \frac{n_j}{N} \right) y_j^2 - \sum_{\substack{j,k \\ j \neq k}} \frac{n_j n_k}{N} y_j y_k .$$

The moment generating function of SST is

$$m_{SST}(t) = E\left(e^{(SST)t} \right)$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi})^p \sigma_1 \sigma_2 \cdots \sigma_p} e^{-\frac{1}{2} \sum_{i=1}^{p} \frac{y_i^2}{\sigma_i^2} + (SST)t} dy_1 \, dy_2 \cdots dy_p$$

$$= \frac{1}{\sqrt{|B|}\ \sigma_1 \sigma_2 \cdots \sigma_p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\sqrt{|B|}}{(\sqrt{2\pi})^p}\, e^{-\frac{1}{2} y' B y}\, dy_1\, dy_2 \cdots dy_p$$

$$= \frac{1}{\sigma_1\, \sigma_2 \cdots \sigma_p \sqrt{|B|}} \quad ,$$

where

$$\sigma_i^2 = \sigma_T^2 + \frac{\sigma^2}{n_i} \qquad (i = 1, 2, \cdots, p)$$

and $|B|$ is the determinant of the matrix

$$B = \begin{pmatrix} \frac{1}{\sigma_1^2} - 2tn_1\left(1 - \frac{n_1}{N}\right) & \frac{2n_1 n_2 t}{N} & \cdots & \frac{2n_1 n_p t}{N} \\ \frac{2n_1 n_2 t}{N} & \frac{1}{\sigma_2^2} - 2tn_2\left(1 - \frac{n_2}{N}\right) & \cdots & \frac{2n_2 n_p t}{N} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{2n_1 n_p t}{N} & \frac{2n_2 n_p t}{N} & \cdots & \frac{1}{\sigma_p^2} - 2tn_p\left(1 - \frac{n_p}{N}\right) \end{pmatrix}$$

If $p=3$,

$$m_{SST}(t) = \Big\{ 1 - \frac{2t}{N}\left[n_1(N-n_1)\sigma_1^2 + n_2(N-n_2)\sigma_2^2 + n_3(N-n_3)\sigma_3^2\right]$$

$$+ \frac{4n_1 n_2 n_3}{N} t^2 \left(\sigma_1^2\sigma_2^2 + \sigma_2^2\sigma_3^2 + \sigma_3^2\sigma_1^2\right)^2 \Big\}^{-1}$$

In order that SST be distributed as $\chi^2 c$, its moment generating function must be

$$(1-2ct)^{-1} \quad ,$$

that is, the above quadratic in t must be a perfect square. It is sufficient if we prove this is not the case when $n_1 = 1$, $n_2 = 2, n_3 = 3$ so that $N = 6$. Then the quadratic becomes

$$1 - \frac{2t}{6}(12\sigma^2 + 22\sigma_\tau^2) + \frac{4t^2}{6}(6\sigma^4 + 22\sigma_\tau^2\sigma^2 + 18\sigma_\tau^4)$$

which is not a perfect square unless $\sigma_\tau = 0$.

Thus we can not hope that SST will be distributed as $\chi^2 c$ when the n_i's are not all equal and σ_τ is not equal to zero.

3.6 Components of Variance for the Type III Model

We shall consider the model

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad ,$$

where $(i = 1,2,\ldots,p)$ and $(j = 1,2,\ldots,n_i)$. The τ_i's come from a finite population of size $P > p$ with mean zero and variance σ_τ^2 defined by

$$\text{(3.61)} \qquad \sigma_\tau^2 = \frac{1}{P-1}\sum_{i=1}^{P}\tau_i^2 = \frac{P}{P-1}E(\tau_i^2) \quad ,$$

and the ε_{ij}'s are $NID(0,\sigma^2)$. Since the τ_i's are no longer independent, we must also consider the covariance between any pair drawn at random. That is, we want to find the $\mathrm{cov}(\tau_i, \tau_{i'}) = E(\tau_i\tau_{i'})$ where $i \neq i'$. It is possible to obtain $P(P-1)$ ordered pairs of the τ_i's so that

$$\text{(3.62)} \qquad E(\tau_i\tau_j) = \sum_{i\neq j}^{P}\frac{(\tau_i\tau_j)}{P(P-1)} \quad .$$

Since

$$\sum_{i=1}^{P}\tau_i = \tau_1 + \tau_2 + \cdots + \tau_P = 0 \quad ,$$

squaring both sides we have

$$\sum_{i=1}^{p} \tau_i^2 + \sum_{i \neq j}^{p} \tau_i \tau_j = 0$$

Therefore, from (3.61)

$$\sum_{i \neq j}^{p} \tau_i \tau_j = -(p-1)\,\sigma_\tau^2 \quad ,$$

and substituting this result in (3.62) gives

$$E(\tau_i \tau_j) = -\frac{\sigma_\tau^2}{p}$$

Once again we shall use the sums of squares obtained in section 3.3. In terms of the original model,

$$(3.63) \qquad \bar{Y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i}(\mu + \tau_i + \varepsilon_{ij}) = \mu + \tau_i + \bar{\varepsilon}_{i.} \quad ,$$

and

$$(3.64) \qquad \bar{Y}_{..} = \frac{1}{N}\sum_{i,j}(\mu + \tau_i + \varepsilon_{ij}) = \mu + \frac{1}{N}\sum_{i=1}^{p} n_i \tau_i + \bar{\varepsilon}_{..} \quad .$$

From equation (3.63) we have

$$E(\bar{Y}_{i.}) = \mu \quad ,$$

$$Var(\bar{Y}_{i.}) = E[(\tau_i + \bar{\varepsilon}_{i.})^2] = E(\tau_i^2) + \frac{\sigma^2}{n_i}$$

$$= \frac{p-1}{p}\sigma_\tau^2 + \frac{\sigma^2}{n_i} = \sigma_i^2 \quad ,$$

and

$$cov(\bar{Y}_{i.}, \bar{Y}_{i'.}) = E[(\tau_i + \bar{\varepsilon}_{i.})(\tau_{i'} + \bar{\varepsilon}_{i'.})] = E(\tau_i \tau_{i'})$$

$$= -\frac{\sigma_\tau^2}{p} = \lambda \quad .$$

Also

$$\bar{\mu}_{\cdot} = \frac{1}{N}\sum_{i=1}^{p} n_i \mu = \mu \ .$$

Therefore the expected treatment sum of squares is

$$E(SST) = E\left\{\sum_{i=1}^{p} n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right\}$$

$$= \sum_{i=1}^{p} n_i\left(1-\frac{n_i}{N}\right)\left[\left(1-\frac{1}{p}\right)\sigma_\tau^2 + \frac{\sigma^2}{n_i} + \frac{\sigma_\tau^2}{p}\right]$$

$$= \left(\sum_{i=1}^{p} n_i - \frac{1}{N}\sum_{i=1}^{p} n_i^2\right)\sigma_\tau^2 + \sum_{i=1}^{p}\left(1-\frac{n_i}{N}\right)\sigma^2$$

$$= \left(N - \frac{1}{N}\sum_{i=1}^{p} n_i^2\right)\sigma_\tau^2 + (p-1)\sigma^2 .$$

Hence the expected value of the mean sum of squares due to treatments is given by

(3.65) $$E(MST) = \frac{E(SST)}{p-1} = \sigma^2 + \frac{1}{p-1}\left(N - \frac{1}{N}\sum_{i=1}^{p} n_i^2\right)\sigma_\tau^2 .$$

Once again

$$SSE = \sum_{i,j}(Y_{ij} - \bar{Y}_{i\cdot})^2 = \sum_{i,j}(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2$$

and hence

$$E(SSE) = (N-p)\,\sigma^2 .$$

Therefore

$$E(MSE) = \frac{E(SSE)}{N-p} = \sigma^2 .$$

It is interesting to note that the expected sum of squares due to treatments is of the same form for both the Type II and Type III models.

The hypothesis we wish to test in this Type III case is that $\sigma_\tau = 0$, which implies that $\tau_i = 0$ (i = 1,2,...,P). Subject to this hypothesis, SST has the same form as in the Type I model and hence SST has a $\chi^2\sigma^2$ distribution with p-1 d.f. and is distributed independently of SSE, which is distributed as $\chi^2\sigma^2$ with N-p d.f. whether the hypothesis holds or not. If we reject the hypothesis that $\sigma_\tau = 0$, we may estimate σ_τ^2 and σ^2 as before by equating the mean squares to their expected values. If we accept the hypothesis that $\sigma_\tau = 0$, we pool the sums of squares and degrees of freedom to estimate σ^2.

3.7 Summary

For all three models we test our hypothesis, $\tau_1 = \tau_2 = \cdots = \tau_p = 0$ for the Type I model and $\sigma_\tau = 0$ for the Type II and Type III models, by the statistic F = MST/MSE with p-1 and N-p degrees of freedom.

The analysis-of-variance table used in all three cases is:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Treatments	p-1	$SST = \sum_{i=1}^{p} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$MST = \frac{SST}{p-1}$	$\sigma^2 + \frac{\sum_{i=1}^{p} n_i \tau_i^2}{p-1}$ (TYPE I) or $\sigma^2 + k\,\sigma_\tau^2$ (TYPE II + III)
Error	N-p	$SSE = \sum_{i,j} (Y_{ij} - \bar{Y}_{i.})^2$	$MSE = \frac{SSE}{N-p}$	σ^2
Total	N-1	$\sum_{i,j} (Y_{ij} - \bar{Y}_{..})^2$		

where

$$k = \frac{1}{N(p-1)} \left(N^2 - \sum_{i=1}^{p} n_i^2 \right) .$$

If our hypothesis is accepted, our estimate of σ^2 is

$$(SST + SSE)/(N-1) \quad ,$$

and, if it is rejected, we estimate σ_τ^2 and σ^2 by solving the equations

$$\sigma^2 + k\,\sigma_\tau^2 = MST \quad , \quad \sigma^2 = MSE \ ,$$

for the Type II and Type III models.

If we have $n_i = n$ $(i = 1,2,\ldots,p)$, so that $N = np$, our analysis-of-variance table for all three models becomes:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Treatments	p-1	$SST = n\sum_{i=1}^{p}(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$MST = \frac{SST}{p-1}$	$\sigma^2 + n\,\sigma_\tau^2$
Error	(n-1)p	$SSE = \sum_{i,j}(Y_{ij} - \bar{Y}_{i.})^2$	$MSE = \frac{SSE}{(n-1)p}$	σ^2
Total	np-1	$\sum_{i,j}(Y_{ij} - \bar{Y}_{..})^2$		

where

$$\bar{Y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} Y_{ij} \ , \qquad \bar{Y}_{..} = \frac{1}{p}\sum_{i=1}^{p} \bar{Y}_{i.}$$

and

$$\sigma_\tau^2 = \sum_{i=1}^{p} \frac{\tau_i^2}{p-1}$$

for the Type I model.

To compute the sums of squares we use the formulas

$$SST = \sum_{i=1}^{p} \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N}$$

and

$$\sum_{i,j} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i,j} Y_{ij}^2 - \frac{Y_{..}^2}{N} \quad ,$$

where

$$Y_{i.} = \sum_{j=1}^{n_i} Y_{ij} \quad , \quad Y_{..} = \sum_{i=1}^{p} Y_{i.} \quad .$$

CHAPTER IV

TWO-WAY CROSSED CLASSIFICATION MODELS

4.1 The Type I Model for Proportional Frequencies

We consider an experiment in which p treatments are applied to q blocks. The ith treatment is applied to the jth block n_{ij} times. Display the n_{ij}'s in a table

		Block				Row Totals
		1	2	...	q	
Treatments	1	n_{11}	n_{12}	...	n_{1q}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2q}	$n_{2.}$
	.	.	.	.	.	.
	p	n_{p1}	n_{p2}	...	n_{pq}	$n_{p.}$
Column Totals		$n_{.1}$	$n_{.2}$	...	$n_{.q}$	N

and let $N = \sum_{i,j}^{p,q} n_{ij}$. We shall assume that the n_{ij}'s in a given row are proportional to the n_{ij}'s in any other row. Thus,

$$n_{ij} = k_i n_{1j} \quad (j = 1, 2, \ldots, q)$$

and

$$n_{i.} = \sum_{j=1}^{q} n_{ij} = k_i \sum_{j=1}^{q} n_{1j} = k_i n_{1.} \quad .$$

Hence

$$n_{ij} = \frac{n_{1j} n_{i.}}{n_{1.}} \quad ,$$

and

$$n_{.j} = \sum_{i=1}^{p} n_{ij} = \frac{n_{1j}}{n_{1.}} \sum_{i=1}^{p} n_{i.} = \frac{n_{1j}}{n_{1.}} N \quad .$$

Therefore

$$n_{ij} = \frac{n_{i.}n_{.j}}{N} \quad . \tag{4.11}$$

Consider the model

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk_{ij}} \tag{4.12}$$
$$(i = 1,2,\ldots,p;\, j = 1,2,\ldots,q;\, k_{ij} = 1,2,\ldots,n_{ij})$$

where the $\varepsilon_{ijk_{ij}}$'s are NID$(0,\sigma^2)$ and the parameters are subject to the conditions

$$\sum_{i,j}^{p,q} n_{ij}\tau_i = \sum_{i=1}^{p} n_{i.}\tau_i = 0 , \tag{4.13}$$

$$\sum_{i,j}^{p,q} n_{ij}\beta_j = \sum_{j=1}^{q} n_{.j}\beta_j = 0 , \tag{4.14}$$

$$\sum_{i=1}^{p} n_{ij}(\tau\beta)_{ij} = \sum_{j=1}^{q} n_{ij}(\tau\beta)_{ij} = 0 . \tag{4.15}$$

Making use of (4.11), we see that the conditions (4.15) are equivalent to

$$\sum_{i=1}^{p} n_{i.}(\tau\beta)_{ij} = \sum_{j=1}^{q} n_{.j}(\tau\beta)_{ij} = 0 . \tag{4.16}$$

To show that the above conditions can be satisfied, denote $E(Y_{ijk_{ij}})$ by ξ_{ij} where

$$\xi_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} \quad .$$

Then, if the conditions hold,

$$\bar{\xi}_{i.} = \frac{1}{n_{i.}}\sum_{j=1}^{q} n_{ij}\,\xi_{ij} = \mu + \tau_i \quad ,$$

$$\bar{\xi}_{.j} = \frac{1}{n_{.j}}\sum_{i=1}^{p} n_{ij}\,\xi_{ij} = \mu + \beta_j \quad ,$$

and

$$\bar{\xi}_{..} = \frac{1}{N}\sum_{i,j}^{p,q} n_{ij}\,\xi_{ij} = \frac{1}{N}\sum_{i=1}^{p} n_{i.}\,\bar{\xi}_{i.} = \mu \quad .$$

Thus, we must define

$$\mu = \bar{\xi}_{..} \;,\quad \tau_i = \bar{\xi}_{i.} - \bar{\xi}_{..} \;,\quad \beta_j = \bar{\xi}_{.j} - \bar{\xi}_{..} \;,$$

$$(\tau\beta)_{ij} = \xi_{ij} - \bar{\xi}_{i.} - \bar{\xi}_{.j} + \bar{\xi}_{..}$$

It may be verified that, when the parameters are defined in the above way, conditions (4.13),(4.14) and (4.15) are satisfied. For example,

$$\sum_{i=1}^{p} n_{i.}\,\tau_i = \sum_{i=1}^{p} n_{i.}\,\bar{\xi}_{i.} - \bar{\xi}_{..}\sum_{i=1}^{p} n_{i.} = N\bar{\xi}_{..} - N\bar{\xi}_{..} = 0 ,$$

and

$$\sum_{i=1}^{p} n_{ij}(\tau\beta)_{ij} = \sum_{i=1}^{p} n_{ij}\,\xi_{ij} - \sum_{i=1}^{p} n_{ij}\,\bar{\xi}_{i.} - \sum_{i=1}^{p} n_{ij}\,\bar{\xi}_{.j} + \sum_{i=1}^{p} n_{ij}\,\bar{\xi}_{..}$$

$$= n_{.j}\,\bar{\xi}_{.j} - \frac{n_{.j}}{N}\sum_{i=1}^{p} n_{i.}\,\bar{\xi}_{i.} - n_{.j}\,\bar{\xi}_{.j} + n_{.j}\,\bar{\xi}_{..}$$

$$= -\frac{n_{.j}}{N}\,N\bar{\xi}_{..} + n_{.j}\,\bar{\xi}_{..} = 0 \;.$$

The equations (4.12) may be put in the form

$$(4.17) \qquad Y_{ijk_{ij}} = \mu + \sum_{i'=1}^{p} U_{i'i}\,\tau_{i'} + \sum_{j'=1}^{q} V_{j'j}\,\beta_{j'} + \sum_{i',j'} W_{i'j'ij}(\tau\beta)_{i'j'} + \varepsilon_{ijk_{ij}} ,$$

where

$$U_{ii'} = \delta_{ii'} \;,\quad V_{jj'} = \delta_{jj'} \;,\quad W_{i'j'ij} = \delta_{ii'}\delta_{jj'} \;.$$

If we order the $Y_{ijk_{ij}}$'s in some way, calling them Y_α ($\alpha = 1,2,\ldots,N$), the equations (4.17) may be put in the vector form

$$(4.18) \qquad Y = \mu + \sum_{i'=1}^{p} U_{i'}\,\tau_{i'} + \sum_{j'=1}^{q} V_{j'}\beta_{j'} + \sum_{i',j'} W_{i'j'}(\tau\beta)_{i'j'} + \varepsilon$$

where the $U_{i'}$'s, $V_{j'}$'s and the $W_{i'j'}$'s are the coefficient vectors of the $\tau_{i'}$'s, $\beta_{j'}$'s and $(\tau\beta)_{i'j'}$'s respectively.

Denoting the elements of $U_{i'}$ by $U_{i'\alpha}$ ($\alpha = 1, 2, \ldots, N$), we define

$$\overline{U}_{i'} = \frac{1}{N}\sum_{\alpha=1}^{N} U_{i'\alpha} = \frac{1}{N}\sum_{i,j,k_{ij}} U_{i'i} = \frac{1}{N}\sum_{i,j} n_{ij}\,\delta_{i'i} = \frac{n_{i'.}}{N} ,$$

and $u_{i'\alpha} = U_{i'\alpha} - \overline{U}_{i'}$, so that

$$0 = \sum_{\alpha=1}^{N} u_{i'\alpha} = \sum_{i,j,k_{ij}} u_{i'i} = \sum_{i,j} n_{ij}\,u_{i'i} = \sum_{i=1}^{p} n_{i.}\,u_{i'i} .$$

In the same way we find

$$\overline{V}_{j'} = \frac{n_{.j'}}{N} , \qquad \sum_{j=1}^{q} n_{.j}\,v_{j'j} = 0 .$$

Define

$$\overline{W}_{i'j'} = \sum_{\alpha=1}^{N}\frac{W_{i'j'\alpha}}{N} = \sum_{i,j,k_{ij}}\frac{W_{i'j'ij}}{N} = \sum_{i,j}\frac{n_{ij}\,\delta_{i'i}\,\delta_{j'j}}{N} = \frac{n_{i'j'}}{N}$$

and

$$w_{i'j'\alpha} = W_{i'j'\alpha} - \overline{W}_{i'j'} .$$

Then

$$\sum_{\alpha=1}^{N} w_{i'j'\alpha} = \sum_{i,j} n_{ij}\,w_{i'j'ij} = 0 .$$

Now denote by $\overline{U}_{i'}, \overline{V}_{j'}, \overline{W}_{i'j'}$ the vectors $\overline{U}_{i'}I, \overline{V}_{j'}I$ and $\overline{W}_{i'j'}I$, respectively, and set

$$u_{i'} = U_{i'} - \overline{U}_{i'} , \quad v_{j'} = V_{j'} - \overline{V}_{j'} , \quad w_{i'j'} = W_{i'j'} - \overline{W}_{i'j'} .$$

We may then write (4.18) in the form

$$(4.19) \quad Y = \mu + \sum_{i'=1}^{p}(u_{i'} + \overline{U}_{i'})\,\tau_{i'} + \sum_{j'=1}^{q}(v_{j'} + \overline{V}_{j'})\,\beta_{j'}$$

$$+ \sum_{i',j'} (\omega_{i'j'} + \overline{W}_{i'j'})(\tau\beta)_{i'j'} + \varepsilon$$

$$= \mu + \sum_{i'=1}^{p} u_{i'} \tau_{i'} + \sum_{j'=1}^{q} v_{j'} \beta_{j'} + \sum_{i',j'}^{p,q} \omega_{i'j'} (\tau\beta)_{i'j'} + \varepsilon$$

since

$$\frac{1}{N} \sum_{i'=1}^{p} n_{i'.} \tau_{i'} = \frac{1}{N} \sum_{j'=1}^{q} n_{.j'} \beta_{j'} = \frac{1}{N} \sum_{i',j'}^{p,q} n_{i'j'} (\tau\beta)_{i'j'} = 0 .$$

If we are to apply the theory of Chapter II, it is necessary that the $u_{i'}$'s, $v_{j'}$'s, and $w_{i'j'}$'s form a linearly independent set of vectors. Unfortunately, this is not the case. Before showing this and remedying the situation, we need certain relations among the vectors. We first show that

$$\sum_{i'=1}^{p} U_{i'} = I$$

This follows since, when we add the elements in row α of these vectors, we have

$$\sum_{i'=1}^{p} U_{i'\alpha} = \sum_{i'=1}^{p} U_{i'i} = \sum_{i'=1}^{p} \delta_{i'i} = 1.$$

A similar proof establishes that

$$\sum_{j'=1}^{q} V_{j'} = \sum_{i',j'}^{p,q} W_{i'j'} = I .$$

Also

$$U_i \cdot I = n_{i.} \;,\; V_j \cdot I = n_{.j} \;,\; W_{ij} \cdot I = n_{ij}$$

and the following multiplication table gives us the values of different dot products formed from our vectors

	U_i	V_j	W_{ij}
$U_{i'}$	$n_{i\cdot}\,\delta_{ii'}$	$n_{i'j}$	$\delta_{ii'}\,n_{ij}$
$V_{j'}$	$n_{ij'}$	$n_{\cdot j}\,\delta_{jj'}$	$\delta_{jj'}\,n_{ij'}$
$W_{i'j'}$	$\delta_{ii'}\,n_{i'j'}$	$\delta_{jj'}\,n_{i'j}$	$\delta_{ii'}\,\delta_{jj'}\,n_{ij}$

Now

$$\sum_{i=1}^{p} u_i = \sum_{i=1}^{p} U_i - \sum_{i=1}^{p} \overline{U}_i = 0$$

since

$$\sum_{i=1}^{p} \overline{U}_i = I \sum_{i=1}^{p} \frac{n_{i\cdot}}{N} = I$$

Hence the u_i's, v_j's and w_{ij}'s do not form a linearly independent set. To meet this difficulty we use the relations (4.13), (4.14), and (4.15) to eliminate $\tau_p, \beta_q, (\tau\beta)_{iq}$ (i = 1,2,...,p), and $(\tau\beta)_{pj}$ (j = 1,2,...,q-1). We have

$$\tau_p = -\frac{1}{n_{p\cdot}} \sum_{i=1}^{p-1} n_{i\cdot}\,\tau_i \quad , \quad \beta_q = -\frac{1}{n_{\cdot q}} \sum_{j=1}^{q-1} n_{\cdot j}\,\beta_j \quad ,$$

$$n_{iq}\,(\tau\beta)_{iq} = -\sum_{j=1}^{q-1} n_{ij}\,(\tau\beta)_{ij} \quad ,$$

$$n_{pj}\,(\tau\beta)_{pj} = -\sum_{i=1}^{p-1} n_{ij}\,(\tau\beta)_{ij} \quad .$$

Hence

$$n_{pq}\,(\tau\beta)_{pq} = -\sum_{i=1}^{p-1} n_{iq}\,(\tau\beta)_{iq} = \sum_{i=1}^{p-1} \sum_{j=1}^{q-1} n_{ij}\,(\tau\beta)_{ij} \ .$$

Therefore

$$\sum_{i'=1}^{p} u_{i'}\tau_{i'} = \sum_{i'=1}^{p-1} u_{i'}\tau_{i'} - \frac{u_p}{n_{p\cdot}} \sum_{i'=1}^{p-1} n_{i'\cdot}\tau_{i'}$$

$$= \sum_{i'=1}^{p-1} \left(u_{i'} - \frac{n_{i'\cdot}}{n_{p\cdot}} u_p\right) \tau_{i'} \quad ,$$

$$\sum_{j'=1}^{q} v_{j'}\beta_{j'} = \sum_{j'=1}^{q-1} v_{j'}\beta_{j'} - \frac{v_q}{n_{\cdot q}} \sum_{j'=1}^{q-1} n_{\cdot j'}\beta_{j'}$$

$$= \sum_{j'=1}^{q-1} \left(v_{j'} - \frac{n_{\cdot j'}}{n_{\cdot q}} v_q\right) \beta_{j'} \quad ,$$

$$\sum_{i',j'}^{p,q} w_{i'j'}(\tau\beta)_{i'j'} = \sum_{i',j'}^{p,q} \frac{w_{i'j'}}{n_{i'j'}} \left[n_{i'j'}(\tau\beta)_{i'j'}\right]$$

$$= \sum_{i'=1}^{p-1} \sum_{j'=1}^{q-1} \left[n_{i'j'}(\tau\beta)_{i'j'}\right] \frac{w_{i'j'}}{n_{i'j'}} + \sum_{i'=1}^{p-1} \left[n_{i'q}(\tau\beta)_{i'q}\right] \frac{w_{i'q}}{n_{i'q}}$$

$$+ \sum_{j'=1}^{q-1} \left[n_{pj'}(\tau\beta)_{pj'}\right] \frac{w_{pj'}}{n_{pj'}} + \left[n_{pq}(\tau\beta)_{pq}\right] \frac{w_{pq}}{n_{pq}}$$

$$= \sum_{i'=1}^{p-1} \sum_{j'=1}^{q-1} \left[n_{i'j'}(\tau\beta)_{i'j'}\right] \frac{w_{i'j'}}{n_{i'j'}} - \sum_{i'=1}^{p-1} \sum_{j'=1}^{q-1} \left[n_{i'j'}(\tau\beta)_{i'j'}\right] \frac{w_{i'q}}{n_{i'q}}$$

$$- \sum_{i'=1}^{p-1} \sum_{j'=1}^{q-1} \left[n_{i'j'}(\tau\beta)_{i'j'}\right] \frac{w_{pj'}}{n_{pj'}} + \sum_{i'=1}^{p-1} \sum_{j'=1}^{q-1} \left[n_{i'j'}(\tau\beta)_{i'j'}\right] \frac{w_{pq}}{n_{pq}}$$

$$= \sum_{i'=1}^{p-1} \sum_{j'=1}^{q-1} \left[n_{i'j'}(\tau\beta)_{i'j'}\right]\left[\frac{w_{i'j'}}{n_{i'j'}} - \frac{w_{i'q}}{n_{i'q}} - \frac{w_{pj'}}{n_{pj'}} + \frac{w_{pq}}{n_{pq}}\right] .$$

We can now write equation (4.19) in the form

$$(4.110)\quad Y = \mu + \sum_{i'=1}^{p-1} \left(u_{i'} - \frac{n_{i'.}}{n_{p.}} u_p\right)\tau_{i'} + \sum_{j'=1}^{q-1} \left(v_{j'} - \frac{n_{.j'}}{n_{.q}} v_q\right)\beta_{j'}$$

$$+ \sum_{i'=1}^{p-1} \sum_{j'=1}^{q-1} n_{i'j'} \left(\frac{w_{i'j'}}{n_{i'j'}} - \frac{w_{i'q}}{n_{i'q}} - \frac{w_{pj'}}{n_{pj'}} + \frac{w_{pq}}{n_{pq}} \right)(\tau\beta)_{i'j'}$$

$$+ \varepsilon \quad .$$

Note that

$$u_{i'} - \frac{n_{i'.}}{n_{p.}} u_p = U_{i'} - \bar{U}_{i'} - \frac{n_{i'.}}{n_{p.}} (U_p - \bar{U}_p)$$

$$= U_{i'} - \frac{n_{i'.}}{n_{p.}} U_p - \frac{n_{i'.}}{N} I + \frac{n_{i'.}}{n_{p.}} \frac{n_{p.}}{N} I = U_{i'} - \frac{n_{i'.}}{n_{p.}} U_p \quad .$$

Similarly

$$v_{j'} - \frac{n_{.j'}}{n_{.q}} v_q = V_{j'} - \frac{n_{.j'}}{n_{.q}} V_q \quad .$$

Also

$$\frac{w_{i'j'}}{n_{i'j'}} - \frac{w_{i'q}}{n_{i'q}} - \frac{w_{pj'}}{n_{pj'}} + \frac{w_{pq}}{n_{pq}} = \frac{1}{n_{i'j'}} (W_{i'j'} - \bar{W}_{i'j'}) - \frac{1}{n_{i'q}} (W_{i'q} - \bar{W}_{i'q})$$

$$- \frac{1}{n_{pj'}} (W_{pj'} - \bar{W}_{pj'}) + \frac{1}{n_{pq}} (W_{pq} - \bar{W}_{pq})$$

$$= \frac{W_{i'j'}}{n_{i'j'}} - \frac{W_{i'q}}{n_{i'q}} - \frac{W_{pj'}}{n_{pj'}} + \frac{W_{pq}}{n_{pq}} - \frac{n_{i'j'}}{n_{i'j'}} I + \frac{n_{i'q}}{n_{i'q}} I + \frac{n_{pj'}}{n_{pj'}} I - \frac{n_{pq}}{n_{pq}} I$$

$$= \frac{W_{i'j'}}{n_{i'j'}} - \frac{W_{i'q}}{n_{i'q}} - \frac{W_{pj'}}{n_{pj'}} + \frac{W_{pq}}{n_{pq}} \quad .$$

Consider the equation

$$\sum_{i'=1}^{p-1} c_{i'} \left(U_{i'} - \frac{n_{i'.}}{n_{p.}} U_p \right) = 0$$

Multiplying the equation by U_i $(i = 1,2,\ldots,p-1)$ we find that $n_{i.} c_i = 0$, since the $U_{i'}$'s are orthogonal vectors. Thus the vectors

$$U_{i'} - \frac{n_{i'.}}{n_{p.}} U_p \qquad (i' = 1,2,\ldots,p-1)$$

are linearly independent. Similarly the

$$V_{j'} - \frac{n_{.j'}}{n_{.q}} V_q \qquad (j' = 1,2,\ldots,q-1)$$

are linearly independent. Consider the equation

$$\sum_{i'=1}^{p-1} \sum_{j'=1}^{q-1} n_{i'j'} c_{i'j'} \left(\frac{W_{i'j'}}{n_{i'j'}} - \frac{W_{i'q}}{n_{i'q}} - \frac{W_{pj'}}{n_{pj'}} + \frac{W_{pq}}{n_{pq}} \right) = 0$$

Multiplying by W_{ij} $(i = 1,2,\ldots,p-1; j = 1,2,\ldots,q-1)$ we find that $n_{ij} c_{ij} = 0$ since the $W_{i'j'}$'s are orthogonal. Therefore these $(p-1)(q-1)$ vectors are linearly independent. Also

$$\left(U_{i'} - \frac{n_{i'.}}{n_{p.}} U_p \right) \cdot \left(V_{j'} - \frac{n_{.j'}}{n_{.q}} V_q \right) = n_{i'j'} - \frac{n_{.j'}}{n_{.q}} n_{i'q} - \frac{n_{i'.}}{n_{p.}} n_{pj'} + \frac{n_{i'.} n_{.j'}}{n_{p.} n_{.q}} n_{pq}$$

$$= n_{i'j'} - \frac{n_{.j'}}{n_{.q}} \frac{n_{i'.} n_{.q}}{N} - \frac{n_{i'.}}{n_{p.}} \frac{n_{p.} n_{.j'}}{N} + \frac{n_{i'.} n_{.j'}}{n_{p.} n_{.q}} \frac{n_{p.} n_{.q}}{N}$$

$$= n_{i'j'} - n_{i'j'} - n_{i'j'} + n_{i'j'} = 0$$

Therefore the two sets of vectors are orthogonal. Next we have

$$U_i \cdot \left(\frac{W_{i'j'}}{n_{i'j'}} - \frac{W_{i'q}}{n_{i'q}} - \frac{W_{pj'}}{n_{pj'}} + \frac{W_{pq}}{n_{pq}} \right)$$

$$= \delta_{ii'} - \delta_{ii'} - \delta_{ip} + \delta_{ip} = 0 \qquad (i, i' = 1, \cdots, p-1 \; ; \; j = 1, \cdots, q-1)$$

and

$$V_j \cdot \left(\frac{W_{i'j'}}{n_{i'j'}} - \frac{W_{i'q}}{n_{i'q}} - \frac{W_{pj'}}{n_{pj'}} + \frac{W_{pq}}{n_{pq}} \right) = 0 \quad .$$

Therefore the third set of vectors is orthogonal to the first two sets. Our model (4.110) now satisfies the conditions required in Chapter II. We shall be interested in testing three hypotheses

H_1: $(\tau\beta)_{ij} = 0$ $\quad (i = 1,2,\ldots,p-1; j = 1,2,\ldots,q-1)$,
H_2: $\tau_i = 0$ $\quad (i = 1,2,\ldots,p-1)$,
H_3: $\beta_j = 0$ $\quad (j = 1,2,\ldots,q-1)$.

Conditions (4.13),(4.14), and (4.15) together with the above hypotheses imply that not only the parameters referred to in a given hypothesis are zero but also all other parameters of the same kind.

To test H_1, we first compute

$$SSE = \sum_{i,j,k_{ij}} \left[Y_{ijk_{ij}} - m - t_i - b_j - (tb)_{ij} \right]^2 ,$$

where m, t_i, b_j and $(tb)_{ij}$ are the least squares estimates of μ, τ_i, β_j and $(\tau\beta)_{ij}$,respectively, and SSE is the minimized value of the residual sum of squares. Next, we compute SSE_1, the corresponding minimum obtained under the assumption that H_1 holds. Then

$$R = \sum_{\alpha=1}^{N} y_\alpha^2 - SSE$$

is the reduction in the sum of squares when all the parameters are used. Also

$$R_1 = \sum_{\alpha=1}^{N} y_\alpha^2 - SSE_1$$

is the reduction due to the parameters left when H_1 is true. Since more parameters are involved in the first case than in the second, $R \geq R_1$, and the additional reduction in the sum of squares due to the $(\tau\beta)_{ij}$'s is

$$SS(TB) = R - R_1 = SSE_1 - SSE = \sum_{i=1}^{p-1} \sum_{j=1}^{q-1} [(tb)'_{ij}]^2,$$

where the $(tb)'_{ij}$'s are the estimates obtained when the orthogonal model of Chapter II is used.

In the same way, SSE_2 and SSE_3 denote the minima obtained subject to H_2 and H_3, respectively, and the reductions in the sum of squares due to the τ_i's and the β_j's are

$$SST = R - R_2 = SSE_2 - SSE = \sum_{i=1}^{p-1} (t'_i)^2,$$

and

$$SSB = R - R_3 = SSE_3 - SSE = \sum_{j=1}^{q-1} (b'_j)^2,$$

respectively. Finally,

$$SSE = \sum_{\alpha=1}^{N} y_\alpha^2 - R$$

$$= \sum_{\alpha=1}^{N} y_\alpha^2 - \sum_{i=1}^{p-1} (t_i')^2 - \sum_{j=1}^{q-1} (b_j')^2 - \sum_{i=1}^{p-1} \sum_{j=1}^{q-1} [(tb)_{ij}']^2 ,$$

so that

$$\sum_{\alpha=1}^{N} y_\alpha^2 = SST + SSB + SS(TB) + SSE .$$

The theory in Chapter II also tells us that, subject to the corresponding hypotheses, SST, SSB, SS(TB), and SSE are independently distributed as $\chi^2\sigma^2$ with p-1, q-1, (p-1)(q-1), and N-pq degrees of freedom, respectively. The hypotheses H_1, H_2, and H_3 are tested by the statistics

$$F_1 = \frac{MS(TB)}{MSE} , \quad F_2 = \frac{MST}{MSE} , \quad F_3 = \frac{MSB}{MSE} ,$$

respectively.

In the next sections methods for the computation of these statistics are developed.

4.2 The Sums of Squares

Our estimates of μ, τ_i, β_j, $(\tau\beta)_{ij}$ are m, t_i, b_j, $(tb)_{ij}$, respectively, where these values minimize

$$SSE = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - m - t_i - b_j - (tb)_{ij})^2$$

subject to the conditions

$$(4.21) \qquad \sum_{i=1}^{p} n_{ij} t_i = \sum_{i=1}^{p} n_{i\cdot} t_i = 0 ,$$

$$(4.22) \qquad \sum_{j=1}^{q} n_{ij} b_j = \sum_{j=1}^{q} n_{\cdot j} b_j = 0 ,$$

$$(4.23) \qquad \sum_{i=1}^{p} n_{ij}(tb)_{ij} = \sum_{i=1}^{p} n_{i\cdot}(tb)_{ij} = 0 \quad ,$$

$$(4.24) \qquad \sum_{j=1}^{q} n_{ij}(tb)_{ij} = \sum_{j=1}^{q} n_{\cdot j}(tb)_{ij} = 0 \quad .$$

By Theorem 3.1, we can ignore conditions (4.21) and (4.22) since their Lagrange multipliers will be zero. Conditions (4.23) and (4.24) will have to be considered in the computation of SSE_2 and SSE_3 but in the computation of SSE they can be avoided by expressing SSE in a different form. We have

$$SSE = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \xi_{ij})^2 ,$$

where $E(Y_{ijk_{ij}}) = \xi_{ij}$. Then

$$\frac{\partial SSE}{\partial \xi_{ij}} = -2 \sum_{k_{ij}=1}^{n_{ij}} (Y_{ijk_{ij}} - \xi_{ij}) = 0 \quad ,$$

and our estimate of ξ_{ij} is $\hat{\xi}_{ij} = \bar{Y}_{ij\cdot}$. Then, by the invariance property of such estimators,[1]

$$m = \bar{\bar{\xi}}_{\cdot\cdot} = \frac{1}{N} \sum_{ij} n_{ij}\hat{\xi}_{ij} = \frac{1}{N} \sum_{i,j,k_{ij}} Y_{ijk_{ij}} = \bar{Y}_{\cdots} \quad ,$$

$$t_i = \bar{\hat{\xi}}_{i\cdot} - m = \frac{1}{n_{i\cdot}} \sum_{j=1}^{q} n_{ij}\bar{Y}_{ij\cdot} - m = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots} \quad ,$$

$$b_j = \bar{\hat{\xi}}_{\cdot j} - m = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots} \quad ,$$

$$(tb)_{ij} = \hat{\xi}_{ij} - \bar{\hat{\xi}}_{i\cdot} - \bar{\hat{\xi}}_{\cdot j} + \bar{\hat{\xi}}_{\cdot\cdot} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots} \quad ,$$

and

$$SSE = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij\cdot})^2 \;.$$

---

[1] Mood, A.M., Introduction to the Theory of Statistics. New York: McGraw-Hill Co., 1950, p. 159.

To obtain SSE_1, we must minimize

$$SSE_1 = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - m - t_i - b_j)^2 \ .$$

Now conditions (4.23) and (4.24) do not apply and, as before we can ignore conditions (4.21) and (4.22). We have

$$-\frac{1}{2}\frac{\partial SSE_1}{\partial m} = Y_{\cdots} - Nm = 0 \ ,$$

$$-\frac{1}{2}\frac{\partial SSE_1}{\partial t_i} = Y_{i\cdot\cdot} - n_{i\cdot}(m + t_i) = 0 \ ,$$

and

$$-\frac{1}{2}\frac{\partial SSE_1}{\partial b_j} = Y_{\cdot j\cdot} - n_{\cdot j}(m + b_j) = 0$$

as our normal equations which have been simplified by the use of conditions (4.21) and (4.22). We conclude that

$$m = \bar{Y}_{\cdots} \ , \quad t_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots} \ , \quad b_j = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots} \ ,$$

and

$$SSE_1 = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2 \ .$$

Then

$$\begin{aligned} SS(TB) &= SSE_1 - SSE \\ &= \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2 - \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij\cdot})^2 \\ &= \sum_{i,j,k_{ij}} (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2 + 2\sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij\cdot})(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots}) \\ &= \sum_{i,j,k_{ij}} (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2 \end{aligned}$$

since the second sum is equal to

$$2\sum_{i,j} n_{ij}(\bar{Y}_{ij\cdot} - \bar{Y}_{ij\cdot})(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot}) = 0 \quad .$$

To determine SSE_2, we minimize

$$SSE_2 = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - m - b_j - (tb)_{ij})^2$$

subject to condition (4.22), which may be ignored, and conditions (4.23) and (4.24). Thus, we must minimize the expression

$$Q = SSE_2 + \sum_{j=1}^{q} c_j \sum_{i=1}^{p} n_{i\cdot}(tb)_{ij} + \sum_{i=1}^{p} d_i \sum_{j=1}^{q} n_{\cdot j}(tb)_{ij} \ .$$

Taking partial derivatives with respect to m and b_j gives the same normal equations as before so

$$m = \bar{Y}\ldots \qquad \text{and} \qquad b_j = \bar{Y}_{\cdot j\cdot} - \bar{Y}\ldots \quad .$$

Then

$$(4.25) \qquad \frac{\partial Q}{\partial (tb)_{ij}} = -2n_{ij}\left[\bar{Y}_{ij\cdot} - m - b_j - (tb)_{ij}\right] + n_{i\cdot}c_j + n_{\cdot j}d_i = 0 \, ,$$

and

$$m + b_j + (tb)_{ij} = \bar{Y}_{ij\cdot} - \frac{1}{2}\,\frac{n_{i\cdot}c_j + n_{\cdot j}d_i}{n_{ij}}$$

$$= \bar{Y}_{ij\cdot} - \frac{N}{2}\left(\frac{c_j}{n_{\cdot j}} + \frac{d_i}{n_{i\cdot}}\right)$$

by (4.11). Thus

$$SSE_2 = \sum_{i,j,k_{ij}} \left[Y_{ijk_{ij}} - \bar{Y}_{ij\cdot} + \frac{N}{2}\left(\frac{c_j}{n_{\cdot j}} + \frac{d_i}{n_{i\cdot}}\right)\right]^2 \ .$$

Summing (4.25) with respect to j,

$$(4.26) \qquad -2n_{i\cdot}\,\bar{Y}_{i\cdot\cdot} + 2n_{i\cdot}\,m + n_{i\cdot}\sum_{j=1}^{q} c_j + N d_i = 0 \quad ,$$

and, summing (4.26) with respect to i,

$$-2N\bar{Y}_{\cdot\cdot\cdot} + 2N\bar{Y}_{\cdot\cdot\cdot} + N\sum_{j=1}^{q} c_j + N\sum_{i=1}^{p} d_i = 0 \quad ,$$

so that

$$(4.27) \qquad \sum_{j=1}^{q} c_j + \sum_{i=1}^{p} d_i = 0 \quad .$$

Next, sum (4.25) with respect to i to obtain

$$\begin{aligned}(4.28) \quad & -2n_{\cdot j}\bar{Y}_{\cdot j\cdot} + 2n_{\cdot j}m + 2n_{\cdot j}b_j + Nc_j + n_{\cdot j}\sum_{i=1}^{p} d_i \\ &= -2n_{\cdot j}(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot}) + Nc_j + n_{\cdot j}\sum_{i=1}^{p} d_i \\ &= Nc_j + n_{\cdot j}\sum_{i=1}^{p} d_i = 0\end{aligned}$$

From equations (4.26) and (4.28), we find

$$\frac{N}{2}\left(\frac{c_j}{n_{\cdot j}} + \frac{d_i}{n_{i\cdot}}\right) = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot} - \frac{1}{2}\left(\sum_{j=1}^{q} c_j + \sum_{i=1}^{p} d_i\right) = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}$$

by (4.27). Thus

$$SSE_2 = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij\cdot} + \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 ,$$

$$m + b_j + (tb)_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} + \bar{Y}_{\cdot\cdot\cdot} \quad ,$$

and

$$(tb)_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot} \quad ,$$

as before. Also

$$\begin{aligned}SST &= SSE_2 - SSE \\ &= \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij\cdot} + \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 - \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij\cdot})^2 \\ &= \sum_{i,j,k_{ij}} (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 = \sum_{i=1}^{p} n_{i\cdot}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 \quad .\end{aligned}$$

In the same way we find

$$SSB = SSE_3 - SSE = \sum_{j=1}^{q} n_{\cdot j} (\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

4.3 Other Models

We still assume that

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk_{ij}} .$$

For the Type II model we assume that the τ_i's , β_j's , $(\tau\beta)_{ij}$'s and $\varepsilon_{ijk_{ij}}$'s are NID with zero means and variances σ_τ^2, σ_β^2, $\sigma_{\tau\beta}^2$, σ^2, respectively. We then have $E(Y_{ijk_{ij}}) = \mu$ and

$$Var(Y_{ijk_{ij}}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2 .$$

For the Type III model we assume that the τ_i's, β_j's , and $(\tau\beta)_{ij}$'s come from finite independent populations of size $P > p$, $Q > q$, and PQ, respectively, with zero means and variances

$$\sigma_\tau^2 = \frac{1}{P-1} \sum_{i=1}^{P} \tau_i^2 , \qquad \sigma_\beta^2 = \frac{1}{Q-1} \sum_{j=1}^{Q} \beta_j^2 ,$$

$$\sigma_{\tau\beta}^2 = \frac{1}{(P-1)(Q-1)} \sum_{i,j}^{P,Q} (\tau\beta)_{ij}^2 = 0 ,$$

The assumption of zero means implies that

$$\sum_{i=1}^{P} \tau_i = 0 , \quad \sum_{j=1}^{Q} \beta_j = 0 , \quad \sum_{i,j}^{P,Q} (\tau\beta)_{ij} = 0 ,$$

and, in addition we assume that

$$\sum_{i=1}^{P} (\tau\beta)_{ij} = \sum_{j=1}^{Q} (\tau\beta)_{ij} = 0 .$$

For the mixed model we may assume that the τ_i's , β_j's, and $(\tau\beta)_{ij}$'s are of any of the types described above. In addition, when the τ_i's , say, are of Type I and the β_j's of Type II, it is sometimes assumed that, corresponding to each β_j, there exists a population of $(\tau\beta)_{ij}$'s consisting of p elements such that

$$\sum_{i=1}^{p} (\tau\beta)_{ij} = 0 \quad , \qquad \sigma_{\tau\beta}^2 = \frac{1}{p-1} \sum_{i=1}^{p} (\tau\beta)_{ij}^2 \quad .$$

If the τ_i's came from a Type III population, we would replace p by P in the above definitions, and if the roles of the τ_i's and β_j's were interchanged, we would interchange i and j and replace p by q. We always assume the $\varepsilon_{ijk_{ij}}$'s are NID$(0,\sigma^2)$.

4.4 The Expected Values of the Sums of Squares

In every case we shall arbitrarily begin with the sums of squares obtained for the Type I model and, since

$$\bar{Y}_{ij.} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \bar{\varepsilon}_{ij.} \quad ,$$

$$\bar{Y}_{i..} = \mu + \tau_i + \bar{\beta}_. + (\overline{\tau\beta})_{i.} + \bar{\varepsilon}_{i..} \quad ,$$

$$\bar{Y}_{.j.} = \mu + \bar{\tau}_. + \beta_j + (\overline{\tau\beta})_{.j} + \bar{\varepsilon}_{.j.} \quad ,$$

$$\bar{Y}_{...} = \mu + \bar{\tau}_. + \bar{\beta}_. + (\overline{\tau\beta})_{..} + \bar{\varepsilon}_{...} \quad ,$$

where

$$\bar{\tau}_. = \frac{1}{N} \sum_{i=1}^{p} n_{i.} \tau_i \quad , \quad \bar{\beta}_. = \frac{1}{N} \sum_{j=1}^{q} n_{.j} \beta_j \quad ,$$

$$(\overline{\tau\beta})_{i.} = \frac{1}{N} \sum_{j=1}^{q} n_{ij} (\tau\beta)_{ij} \quad , \quad (\overline{\tau\beta})_{.j} = \frac{1}{N} \sum_{i=1}^{p} n_{i.} (\tau\beta)_{ij} \quad ,$$

$$(\overline{\tau\beta})_{..} = \frac{1}{N} \sum_{i,j}^{p,q} n_{ij} (\tau\beta)_{ij} \quad ,$$

$$SST = \sum_{i=1}^{p} n_{i\cdot} (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$= \sum_{i=1}^{p} n_{i\cdot} [\tau_i - \bar{\tau}_\cdot + (\overline{\tau\beta})_{i\cdot} - (\overline{\tau\beta})_{\cdot\cdot} + \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot\cdot\cdot}]^2 ,$$

$$SSB = \sum_{j=1}^{q} n_{\cdot j} (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$= \sum_{j=1}^{q} n_{\cdot j} [\beta_j - \bar{\beta}_\cdot + (\overline{\tau\beta})_{\cdot j} - (\overline{\tau\beta})_{\cdot\cdot} + \bar{\varepsilon}_{\cdot j\cdot} - \bar{\varepsilon}_{\cdot\cdot\cdot}]^2 ,$$

$$SS(TB) = \sum_{i,j}^{p,q} n_{ij} (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$$

$$= \sum_{i,j}^{p,q} n_{ij} [(\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} - (\overline{\tau\beta})_{\cdot j} + (\overline{\tau\beta})_{\cdot\cdot} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}_{\cdot\cdot\cdot}]^2 ,$$

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij\cdot})^2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (\varepsilon_{ijk_{ij}} - \bar{\varepsilon}_{ij\cdot})^2 .$$

We wish to use theorem 3.2 to evaluate the expected values of the above sums of squares, and, to do so, we shall need the variances and covariances of the following sets of variables,

$$\tau_i + (\overline{\tau\beta})_{i\cdot} + \bar{\varepsilon}_{i\cdot\cdot} ,$$

$$\beta_j + (\overline{\tau\beta})_{\cdot j} + \bar{\varepsilon}_{\cdot j\cdot} ,$$

$$(\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} .$$

To obtain these, we must compute

$$E(\tau_i^2),\ E(\tau_i \tau_{i'}),\ E(\beta_j^2),\ E(\beta_j \beta_{j'}),\ E[(\overline{\tau\beta})_{i\cdot}^2],\ E[(\overline{\tau\beta})_{i\cdot} (\overline{\tau\beta})_{i'\cdot}] ,$$

$$E[(\overline{\tau\beta})_{.j}^2],\ E[(\overline{\tau\beta})_{.j}(\overline{\tau\beta})_{.j'}],\ E[(\tau\beta)_{ij}(\overline{\tau\beta})_{i.}],\ E[(\tau\beta)_{ij}^2],$$

$$E[(\tau\beta)_{ij}(\tau\beta)_{i'j}],\ \text{and}\ E[(\tau\beta)_{ij}(\overline{\tau\beta})_{i'.}],$$

in a form which will be valid regardless of the nature of the populations used.

If τ_i comes from a Type II population, $E(\tau_i^2) = \sigma_\tau^2$, and if it comes from a Type III population,

$$E(\tau_i^2) = \frac{1}{P}\sum_{i=1}^{P}\tau_i^2 = \left(1-\frac{1}{P}\right)\sigma_\tau^2 ,$$

which gives the formula for the Type II case if we let $P \to \infty$. For the Type II case, $E(\tau_i\tau_{i'}) = 0$, and for the Type III case

$$E(\tau_i\tau_{i'}) = \sum_{i\neq i'}^{P}\frac{\tau_i\tau_{i'}}{P(P-1)} = -\sum_{i=1}^{P}\frac{\tau_i^2}{P(P-1)} = -\frac{\sigma_\tau^2}{P}$$

since

$$0 = \sum_{i,i'}^{P}\tau_i\tau_{i'} = \sum_{i=1}^{P}\tau_i^2 + \sum_{i\neq i'}^{P}\tau_i\tau_{i'} .$$

Thus the formula for the Type II case is again included in the Type III case. Similarly, for the β_j's,

$$E(\beta_j^2) = \left(1-\frac{1}{Q}\right)\sigma_\beta^2 ,\quad E(\beta_j\beta_{j'}) = -\frac{\sigma_\beta^2}{Q} .$$

Turning to the $(\tau\beta)_{ij}$'s, for the Type III case,

$$E[(\tau\beta)_{ij}^2] = \sum_{i,j}^{P,Q}\frac{(\tau\beta)_{ij}^2}{PQ} = \left(1-\frac{1}{P}\right)\left(1-\frac{1}{Q}\right)\sigma_{\tau\beta}^2 ,$$

$$E[(\tau\beta)_{ij}(\tau\beta)_{i'j}] = \sum_{\substack{i\neq i'\\ j}}^{P,Q}\frac{(\tau\beta)_{ij}(\tau\beta)_{i'j}}{P(P-1)Q} = -\sum_{i,j}^{P,Q}\frac{(\tau\beta)_{ij}^2}{P(P-1)Q}$$

$$= -\frac{(P-1)(Q-1)\,\sigma_{\tau\beta}^2}{P(P-1)Q} = -\frac{1}{P}\left(1-\frac{1}{Q}\right)\sigma_{\tau\beta}^2 \;,$$

since

$$0 = \sum_{i,i'}^{P}(\tau\beta)_{ij}(\tau\beta)_{i'j} = \sum_{i=i'}^{P}(\tau\beta)_{ij}^2 + \sum_{i\neq i'}^{P}(\tau\beta)_{ij}(\tau\beta)_{i'j} \;,$$

$$E[(\tau\beta)_{ij}(\tau\beta)_{ij'}] = -\frac{1}{Q}\left(1-\frac{1}{P}\right)\sigma_{\tau\beta}^2 \;,$$

$$E[(\tau\beta)_{ij}(\tau\beta)_{i'j'}] = \frac{\sum_{i\neq i', j\neq j'}^{P,Q}(\tau\beta)_{ij}(\tau\beta)_{i'j'}}{P(P-1)Q(Q-1)} = \frac{\sum_{i,j}^{P,Q}(\tau\beta)_{ij}^2}{P(P-1)Q(Q-1)} = \frac{\sigma_{\tau\beta}^2}{PQ} \;,$$

since

$$0 = \sum_{i,i',j,j'}^{P,Q}(\tau\beta)_{ij}(\tau\beta)_{i'j'} = \sum_{i,j}^{P,Q}(\tau\beta)_{ij}^2 + \sum_{i\neq i',j}^{P,Q}(\tau\beta)_{ij}(\tau\beta)_{i'j}$$

$$+ \sum_{i,j\neq j'}^{P,Q}(\tau\beta)_{ij}(\tau\beta)_{ij'} + \sum_{i\neq i',j\neq j'}^{P,Q}(\tau\beta)_{ij}(\tau\beta)_{i'j'}$$

$$= \sum_{i,j}^{P,Q}(\tau\beta)_{ij}^2 - \sum_{i,j}^{P,Q}(\tau\beta)_{ij}^2 - \sum_{i,j}^{P,Q}(\tau\beta)_{ij}^2 + \sum_{i\neq i',j\neq j'}^{P,Q}(\tau\beta)_{ij}(\tau\beta)_{i'j'} \;,$$

and these results may be summarized in the single formula

$$E[(\tau\beta)_{ij}(\tau\beta)_{i'j'}] = \left(\delta_{ii'} - \frac{1}{P}\right)\left(\delta_{jj'} - \frac{1}{Q}\right)\sigma_{\tau\beta}^2 \;.$$

If we let $P \rightarrow \infty$, $Q \rightarrow \infty$, the above formula gives us the correct result for the Type II case since then

$$E[(\tau\beta)_{ij}(\tau\beta)_{i'j'}] = \delta_{ii'}\,\delta_{jj'}\,\sigma_{\tau\beta}^2 \;.$$

In the case of the mixed model where each value of j gives us a different independent population of size $P \geq p$ with

$$\sum_{i=1}^{P}(\tau\beta)_{ij} = 0 \quad , \quad \sigma^2_{\tau\beta} = \frac{1}{P-1}\sum_{i=1}^{P}(\tau\beta)^2_{ij} \quad ,$$

$$E[(\tau\beta)^2_{ij}] = \sum_{i=1}^{P}\frac{(\tau\beta)^2_{ij}}{P} = \left(1-\frac{1}{P}\right)\sigma^2_{\tau\beta} \quad ,$$

$$E[(\tau\beta)_{ij}(\tau\beta)_{i'j}] = -\frac{\sigma^2_{\tau\beta}}{P} \quad ,$$

$$E[(\tau\beta)_{ij}(\tau\beta)_{ij'}] = E[(\tau\beta)_{ij}(\tau\beta)_{i'j'}] = 0 \quad ,$$

for $j \neq j'$, and all these results are included in the results for the Type III case if we let $Q \rightarrow \infty$.

Next we work out the expected values of expressions involving $(\overline{\tau\beta})_{i.}$ and $(\overline{\tau\beta})_{.j}$. There is no problem when the $(\tau\beta)_{ij}$'s come from a Type I population since then these terms are equal to zero. For all other types of populations we have

$$E[(\tau\beta)_{ij}(\overline{\tau\beta})_{i'.}] = \frac{1}{N}\sum_{j'=1}^{q} n_{.j'}\, E[(\tau\beta)_{ij}(\tau\beta)_{i'j'}]$$

$$= \frac{1}{N}\sum_{j'=1}^{q} n_{.j'}\left(\delta_{ii'}-\frac{1}{P}\right)\left(\delta_{jj'}-\frac{1}{Q}\right)\sigma^2_{\tau\beta} = \left(\delta_{ii'}-\frac{1}{P}\right)\frac{1}{N}\left(n_{.j}-\frac{N}{Q}\right)\sigma^2_{\tau\beta} ,$$

$$E[(\overline{\tau\beta})_{i.}(\overline{\tau\beta})_{i'.}] = \frac{1}{N}\, E\sum_{j=1}^{q} n_{.j}(\tau\beta)_{ij}(\overline{\tau\beta})_{i'.}$$

$$= \left(\delta_{ii'}-\frac{1}{P}\right)\frac{1}{N^2}\left[\sum_{j=1}^{q} n^2_{.j} - \frac{N^2}{Q}\right]\sigma^2_{\tau\beta} \quad ,$$

and similarly

$$E[(\overline{\tau\beta})_{.j}(\overline{\tau\beta})_{.j'}] = \left(\delta_{jj'}-\frac{1}{Q}\right)\frac{1}{N^2}\left[\sum_{i=1}^{p} n^2_{i.} - \frac{N^2}{P}\right]\sigma^2_{\tau\beta} \quad .$$

To evaluate E(SST), we let

$$y_i = \tau_i + (\overline{\tau\beta})_{i\cdot} + \bar{\varepsilon}_{i\cdot\cdot} \quad ,$$

$$\bar{y}_\cdot = \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}\, y_i = \bar{\tau}_\cdot + \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}\frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}(\tau\beta)_{ij} + \bar{\varepsilon}_{\cdots}$$

$$= \bar{\tau}_\cdot + \frac{1}{N}\sum_{i,j}^{p,q} n_{ij}(\tau\beta)_{ij} + \bar{\varepsilon}_{\cdots} = \bar{\tau}_\cdot + (\overline{\tau\beta})_{\cdot\cdot} + \bar{\varepsilon}_{\cdots} \, ,$$

so that

$$SST = \sum_{i=1}^{p} n_{i\cdot}(y_i - \bar{y}_\cdot)^2 ,$$

$$\mu_i = E(y_i) = E(\tau_i) + \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}\, E[(\tau\beta)_{ij}] = E(\tau_i) = (1-\delta_\tau)\tau_i \, ,$$

$$\bar{\mu}_\cdot = \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}\,\mu_i = 0 \quad ,$$

where $\delta_\tau = 0$ if the τ_i's come from a Type I population and $\delta_\tau = 1$ otherwise. Also

$$\sigma_i^2 = Var(y_i) = E[\tau_i - E(\tau_i) + (\overline{\tau\beta})_{i\cdot} + \bar{\varepsilon}_{i\cdot\cdot}]^2$$

$$= E[\tau_i - E(\tau_i)]^2 + E[(\overline{\tau\beta})^2_{i\cdot}] + E(\bar{\varepsilon}^2_{i\cdot\cdot})$$

$$= \delta_\tau\left(1-\frac{1}{P}\right)\sigma_\tau^2 + \delta_{\tau\beta}\left(1-\frac{1}{P}\right)\frac{1}{N^2}\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_{\tau\beta}^2 + \frac{\sigma^2}{n_{i\cdot}} \quad ,$$

where $\delta_{\tau\beta} = 0$ if the $(\tau\beta)_{ij}$'s are from a Type I population and $\delta_{\tau\beta} = 1$ otherwise. Next

$$\lambda = cov(y_i, y_{i'}) = E\{[\tau_i - E(\tau_i)][\tau_{i'} - E(\tau_{i'})]\} + E[(\overline{\tau\beta})_{i\cdot}(\overline{\tau\beta})_{i'\cdot}]$$

$$= -\delta_\tau\frac{\sigma_\tau^2}{P} - \delta_{\tau\beta}\frac{1}{PN^2}\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_{\tau\beta}^2 \, ,$$

$$\sigma_i^2 - \lambda = \delta_\tau \sigma_\tau^2 + \delta_{\tau\beta} \frac{1}{N^2}\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_{\tau\beta}^2 + \frac{\sigma^2}{n_{i\cdot}} ,$$

$$E(SST) = (1-\delta_\tau)\sum_{i=1}^{p} n_{i\cdot}\tau_i^2 + \sum_{i=1}^{p} n_{i\cdot}\left(1-\frac{n_{i\cdot}}{N}\right)\left\{\delta_\tau \sigma_\tau^2 + \delta_{\tau\beta}\frac{1}{N^2}\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_{\tau\beta}^2 + \frac{\sigma^2}{n_{i\cdot}}\right\}$$

$$= (p-1)\sigma^2 + \delta_{\tau\beta}\frac{1}{N^2}\left[N - \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}^2\right]\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_{\tau\beta}^2$$

$$+ \delta_\tau\left[N - \sum_{i=1}^{p}\frac{n_{i\cdot}^2}{N}\right]\sigma_\tau^2 + (1-\delta_\tau)\sum_{i=1}^{p} n_{i\cdot}\tau_i^2$$

and

$$E(MST) = \sigma^2 + \delta_{\tau\beta}\frac{1}{(p-1)N^2}\left[N - \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}^2\right]\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_{\tau\beta}^2$$

$$+ \frac{\delta_\tau}{p-1}\left[N - \sum_{i=1}^{p}\frac{n_{i\cdot}^2}{N}\right]\sigma_\tau^2 + \frac{(1-\delta_\tau)}{p-1}\sum_{i=1}^{p} n_{i\cdot}\tau_i^2 .$$

Similarly

$$E(MSB) = \sigma^2 + \frac{\delta_{\tau\beta}}{(q-1)N^2}\left[N - \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}^2\right]\left[\sum_{i=1}^{p} n_{i\cdot}^2 - \frac{N^2}{P}\right]\sigma_{\tau\beta}^2$$

$$+ \frac{\delta_\beta}{q-1}\left[N - \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}^2\right]\sigma_\beta^2 + \frac{(1-\delta_\beta)}{q-1}\sum_{j=1}^{q} n_{\cdot j}\beta_j^2 .$$

To evaluate E[SS(TB)] , we let

$$y_i = (\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} + \varepsilon_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} ,$$

$$\bar{y}_\cdot = \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot} y_i = (\overline{\tau\beta})_{\cdot j} - (\overline{\tau\beta})_{\cdot\cdot} + \bar{\varepsilon}_{\cdot j\cdot} - \bar{\varepsilon}_{\cdots} ,$$

so that

$$SS(TB) = \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}\sum_{i=1}^{p} n_{i\cdot}(y_i - \bar{y}_\cdot)^2 ,$$

$$\mu_i = E(y_i) = (1-\delta_{\tau\beta})(\tau\beta)_{ij} \quad , \quad \bar{\mu}_{\cdot} = \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}\mu_i = 0 \, ,$$

$$\sigma_i^2 = Var(y_i) = E\left\{ (\tau\beta)_{ij} - (1-\delta_{\tau\beta})(\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} \right\}^2$$

$$= E\left\{ \delta_{\tau\beta}(\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} \right\}^2$$

$$= \delta_{\tau\beta}\, E\left[(\tau\beta)_{ij}^2\right] + E\left[(\overline{\tau\beta})_{i\cdot}^2\right] - 2\delta_{\tau\beta}\, E\left[(\tau\beta)_{ij}(\overline{\tau\beta})_{i\cdot}\right]$$

$$+ E(\bar{\varepsilon}_{ij\cdot}^2) + E(\bar{\varepsilon}_{i\cdot\cdot}^2) - 2\,E(\bar{\varepsilon}_{ij\cdot}\,\bar{\varepsilon}_{\cdots})$$

$$= \delta_{\tau\beta}\left\{ \left(1-\frac{1}{P}\right)\left(1-\frac{1}{Q}\right) + \left(1-\frac{1}{P}\right)\frac{1}{N^2}\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\right.$$

$$\left. -2\left(1-\frac{1}{P}\right)\frac{1}{N}\left(n_{\cdot j} - \frac{N}{Q}\right)\right\}\sigma_{\tau\beta}^2 + \frac{\sigma^2}{n_{ij}} + \frac{\sigma^2}{n_{i\cdot}} - \frac{2}{N}\, E\left[\sum_{j'=1}^{q} n_{\cdot j'}\,\bar{\varepsilon}_{ij\cdot}\,\bar{\varepsilon}_{ij'\cdot}\right]$$

$$= \delta_{\tau\beta}\left(1-\frac{1}{P}\right)\left\{1 - \frac{1}{Q} + \frac{1}{N^2}\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{1}{Q} - \frac{2n_{\cdot j}}{N} + \frac{2}{Q}\right\}\sigma_{\tau\beta}^2$$

$$+ \frac{\sigma^2}{n_{ij}} + \frac{\sigma^2}{n_{i\cdot}} - \frac{2n_{\cdot j}}{N}\,\frac{\sigma^2}{n_{\cdot j}}$$

$$= \delta_{\tau\beta}\left(1-\frac{1}{P}\right)\left\{1 - \frac{2n_{\cdot j}}{N} + \frac{1}{N^2}\sum_{j=1}^{q} n_{\cdot j}^2\right\}\sigma_{\tau\beta}^2 + \frac{\sigma^2}{n_{\cdot j}} - \frac{\sigma^2}{n_{i\cdot}} \, ,$$

$$\lambda = cov(y_i\, y_{i'}) = E\left\{\left[\delta_{\tau\beta}(\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot}\right]\left[\delta_{\tau\beta}(\tau\beta)_{i'j}\right.\right.$$

$$\left.\left. - (\overline{\tau\beta})_{i'\cdot} + \bar{\varepsilon}_{i'j\cdot} - \bar{\varepsilon}_{i'\cdot\cdot}\right]\right\}$$

$$= \delta_{\tau\beta}\, E\left[(\tau\beta)_{ij}(\tau\beta)_{i'j}\right] - 2\delta_{\tau\beta}\, E\left[(\tau\beta)_{ij}(\overline{\tau\beta})_{i'\cdot}\right] + E\left[(\overline{\tau\beta})_{i\cdot}(\overline{\tau\beta})_{i'\cdot}\right]$$

$$= \delta_{\tau\beta}\left\{-\frac{1}{P}\left(1-\frac{1}{Q}\right) + \frac{2}{PN}\left(n_{\cdot j} - \frac{N}{Q}\right) - \frac{1}{PN^2}\left(\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right)\right\}\sigma_{\tau\beta}^2$$

$$= \frac{1}{P}\,\delta_{\tau\beta}\left\{-1 + \frac{1}{Q} + \frac{2n_{\cdot j}}{N} - \frac{2}{Q} - \frac{1}{N^2}\sum_{j=1}^{q} n_{\cdot j}^2 + \frac{1}{Q}\right\}\sigma_{\tau\beta}^2$$

$$= -\frac{1}{p}\,\delta_{\tau\beta}\left\{1 - \frac{2n_{\cdot j}}{N} + \frac{1}{N^2}\sum_{j=1}^{q} n_{\cdot j}^2\right\}\sigma_{\tau\beta}^2 \qquad ,$$

$$\sigma_i^2 - \lambda = \delta_{\tau\beta}\left\{1 - \frac{2n_{\cdot j}}{N} + \frac{1}{N^2}\sum_{j=1}^{q} n_{\cdot j}^2\right\}\sigma_{\tau\beta}^2 + \frac{\sigma^2}{n_{\cdot j}} - \frac{\sigma^2}{n_{i\cdot}} \quad ,$$

$$E[SS(TB)] = \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}\left\{\sum_{i=1}^{p} n_{i\cdot}(1-\delta_{\tau\beta})(\tau\beta)_{ij}^2 + \sum_{i=1}^{p} n_{i\cdot}\left(1-\frac{n_{i\cdot}}{N}\right)\left(\delta_{\tau\beta}\left[1 - \frac{2n_{\cdot j}}{N} + \frac{1}{N^2}\sum_{j=1}^{q} n_{\cdot j}^2\right]\sigma_{\tau\beta}^2 + \frac{\sigma^2}{n_{\cdot j}} - \frac{\sigma^2}{n_{i\cdot}}\right)\right\}$$

$$= (1-\delta_{\tau\beta})\sum_{i,j}^{p,q} n_{ij}(\tau\beta)_{ij}^2 + \delta_{\tau\beta}\frac{1}{N}\left[N - \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}^2\right]\left[N - \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}^2\right]\sigma_{\tau\beta}^2 + q(p-1)\sigma^2 - (p-1)\sigma^2 ,$$

and

$$E[MS(TB)] = \sigma^2 + \frac{(1-\delta_{\tau\beta})\sum_{i,j}^{p,q} n_{ij}(\tau\beta)_{ij}^2}{(p-1)(q-1)} + \frac{\delta_{\tau\beta}\left[N - \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}^2\right]\left[N - \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}^2\right]\sigma_{\tau\beta}^2}{(p-1)(q-1)N} .$$

Finally, by the theory of Chapter II, we know that SSE is distributed as $\chi^2\sigma^2$ with N-pq d.f. . Hence

$$E\left(\frac{SSE}{\sigma^2}\right) = N-pq \quad \text{and} \quad E(MSE) = \sigma^2 .$$

These results are summarized in table 4.1 where

$$a = N - \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}^2 \quad , \quad b = N - \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}^2 \quad ,$$

δ_τ , δ_β , $\delta_{\tau\beta}$, are zero if the τ_i's , β_j's , $(\tau\beta)_{ij}$'s come from a Type I population and are one otherwise.

TABLE 4.1

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Treatments	p-1	$SST = \sum_{i=1}^{p} n_{i.}(\overline{Y}_{i..} - \overline{Y}_{...})^2$	$MST = \frac{SST}{p-1}$	$\sigma^2 + \delta_{\tau\beta}\, a \left[\frac{\sum_{j=1}^{q} n_{.j}^2 - \frac{N^2}{Q}}{(p-1)N^2} \right] \sigma_{\tau\beta}^2 + \frac{\delta_\tau\, a\, \sigma_\tau^2}{p-1} + (1-\delta_\tau) \sum_{i=1}^{p} \frac{n_{i.}\, \tau_i^2}{p-1}$
Blocks	q-1	$SSB = \sum_{j=1}^{q} n_{.j}(\overline{Y}_{.j.} - \overline{Y}_{...})^2$	$MSB = \frac{SSB}{q-1}$	$\sigma^2 + \delta_{\tau\beta}\, b \left[\frac{\sum_{i=1}^{p} n_{i.}^2 - \frac{N^2}{P}}{(q-1)N^2} \right] \sigma_{\tau\beta}^2 + \frac{\delta_\beta\, b\, \sigma_\beta^2}{q-1} + (1-\delta_\beta) \sum_{j=1}^{q} \frac{n_{.j}}{q-1} \beta_j^2$
Interaction	(p-1)(q-1)	$SS(TB) = \sum_{i,j}^{p,q} n_{ij}(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2$	$MS(TB) = \frac{SS(TB)}{(p-1)(q-1)}$	$\sigma^2 + \frac{\delta_{\tau\beta}\, a\, b}{(p-1)(q-1)N} \sigma_{\tau\beta}^2 + (1-\delta_{\tau\beta}) \sum_{i,j}^{p,q} \frac{n_{ij}\, (\tau\beta)_{ij}^2}{(p-1)(q-1)}$
Error	N-pq	$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \overline{Y}_{ij.})^2$	$MSE = \frac{SSE}{N-pq}$	σ^2
Total	N-1	$\sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \overline{Y}_{...})^2$		

The following formulas are more convenient for computation:

$$SST = \sum_{i=1}^{p} n_{i\cdot}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 = \sum_{i=1}^{p} n_{i\cdot}\bar{Y}_{i\cdot\cdot}^2 - N\bar{Y}_{\cdot\cdot\cdot}^2 = \sum_{i=1}^{p} \frac{Y_{i\cdot\cdot}^2}{n_{i\cdot}} - \frac{Y_{\cdot\cdot\cdot}^2}{N},$$

$$SSB = \sum_{j=1}^{q} \frac{Y_{\cdot j\cdot}^2}{n_{\cdot j}} - \frac{Y_{\cdot\cdot\cdot}^2}{N},$$

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij\cdot})^2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} Y_{ijk_{ij}}^2 - \sum_{i,j}^{p,q} \frac{Y_{ij\cdot}^2}{n_{ij}},$$

$$\sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{\cdot\cdot\cdot})^2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} Y_{ijk_{ij}}^2 - \frac{Y_{\cdot\cdot\cdot}^2}{N},$$

while SS(TB) may be obtained by subtraction.

If $n_{ij} = 1$ $(i = 1,2,\ldots,p; j = 1,2,\ldots,q)$, $\bar{Y}_{ij\cdot} = Y_{ijk_{ij}}$, SSE = 0, and it is impossible to carry out any of the F tests.

If $n_{ij} = n$, $n_{i\cdot} = qn$, $n_{\cdot j} = pn$, $N = pqn$,

$$a = \frac{npq - pq^2n^2}{pqn} = qn(p-1), \quad b = pn(q-1),$$

$$\frac{1}{N^2}\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right] = \frac{1}{N^2}\left[p^2qn^2 - \frac{p^2q^2n^2}{Q}\right] = \frac{pn}{N}\left(1 - \frac{q}{Q}\right),$$

$$\frac{1}{N^2}\left[\sum_{i=1}^{p} n_{i\cdot}^2 - \frac{N^2}{P}\right] = \frac{qn}{N}\left(1 - \frac{p}{P}\right),$$

$$\frac{\sum_{i=1}^{p} n_{i\cdot}\tau_i^2}{p-1} = \frac{qn\sum_{i=1}^{p}\tau_i^2}{p-1} \quad \text{and} \quad \frac{\sum_{j=1}^{q} n_{\cdot j}\beta_j^2}{q-1} = \frac{pn\sum_{j=1}^{q}\beta_j^2}{q-1}.$$

If we define

$$\sigma_\tau^2 = \frac{\sum_{i=1}^{p}\tau_i^2}{p-1}, \quad \sigma_\beta^2 = \frac{\sum_{j=1}^{q}\beta_j^2}{q-1}, \quad \sigma_{\tau\beta}^2 = \frac{\sum_{i,j}^{p,q}(\tau\beta)_{ij}^2}{(p-1)(q-1)}$$

TABLE 4.2

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Treatments	p-1	$SST = qn \sum_{i=1}^{p} (\bar{Y}_{i..} - \bar{Y}_{...})^2$	$MST = \frac{SST}{p-1}$	$\sigma^2 + n(1 - \frac{q}{Q}) \sigma_{\tau\beta}^2 + qn \sigma_\tau^2$
Blocks	q-1	$SSB = pn \sum_{j=1}^{q} (\bar{Y}_{.j.} - \bar{Y}_{...})^2$	$MSB = \frac{SSB}{q-1}$	$\sigma^2 + n(1 - \frac{p}{P}) \sigma_{\tau\beta}^2 + pn \sigma_\beta^2$
Interaction	(p-1)(q-1)	$SS(TB) = n \sum_{i,j}^{p,q} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$	$MS(TB) = \frac{SS(TB)}{(p-1)(q-1)}$	$\sigma^2 + n \sigma_{\tau\beta}^2$
Error	N-pq	$SSE = \sum_{i,j,k}^{p,q,n} (Y_{ijk} - \bar{Y}_{ij.})^2$	$MSE = \frac{SSE}{N-pq}$	σ^2
Total	N-1	$\sum_{i,j,k}^{p,q,n} (Y_{ijk} - \bar{Y}_{...})^2$		

when $n_{ij} = n$ for the Type I populations, we obtain table 4.2, where we omit $\delta_{\tau\beta}$ since $1 - \frac{q}{Q} = 0$ when $Q = q$.

4.5 Models with No Interaction

In this case, the Type I model is

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_j + \varepsilon_{ijk_{ij}} \quad .$$

We then find as in section 4.2 that

$$m = \overline{Y}\ldots \, , \quad t_i = \overline{Y}_{i..} - \overline{Y}\ldots \, , \quad b_j = \overline{Y}_{.j.} - \overline{Y}\ldots \, ,$$

and

$$SSE_1 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} \left(Y_{ijk_{ij}} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}\ldots \right)^2$$

of that section plays the role of SSE. We also saw in section 4.2 that

$$SSE_1 = SSE + SS(TB)$$

so, if we had accepted

$$H_1 \colon (\tau\beta)_{ij} = 0$$

and decided to change models in midstream, all that would be necessary to obtain SSE_1 would be to <u>pool</u> the interaction and error sum of squares. To test $H_2 \colon \tau_i = 0$, we minimize

$$SSE_2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} \left(Y_{ijk_{ij}} - m - b_j \right)^2$$

where condition (4.21) may be ignored, and find

$$m = \overline{Y}\ldots \, , \quad b_j = \overline{Y}_{.j.} - \overline{Y}\ldots \, ,$$

so that

$$SSE_2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} \left(Y_{ijk_{ij}} - \overline{Y}_{.j.} \right)^2 \quad .$$

Then

$$SST = SSE_2 - SSE_1$$

$$= \sum_{i,j,k_{ij}}^{p,q,n_{ij}} \left[(Y_{ijk_{ij}} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) + (\bar{Y}_{i..} - \bar{Y}_{...}) \right]^2$$

$$- \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

$$= \sum_{i=1}^{p} n_{i.} (\bar{Y}_{i..} - \bar{Y}_{...})^2 + 2 \sum_{i=1}^{p} (\bar{Y}_{i..} - \bar{Y}_{...}) \sum_{j,k_{ij}}^{q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$$

$$= \sum_{i=1}^{p} n_{i.} (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

since

$$\sum_{j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) = n_{i.} \bar{Y}_{i..} - n_{i.} \bar{Y}_{i..} - \frac{n_{i.}}{N} \sum_{j=1}^{q} n_{.j} \bar{Y}_{.j.} + n_{i.} \bar{Y}_{...}$$

$$= n_{i.} \bar{Y}_{i..} - n_{i.} \bar{Y}_{i..} - n_{i.} \bar{Y}_{...} + n_{i.} \bar{Y}_{...} = 0 .$$

Similarly,

$$SSB = \sum_{i=1}^{p} n_{.j} (\bar{Y}_{.j.} - \bar{Y}_{...})^2 ,$$

and we see that the formulas for SST and SSB are the same as in section 4.4, and they have, as before, p-1 and q-1 degrees of freedom. The degrees of freedom associated with SSE_1 are

$$N-1-(p-1)-(q-1) = N-p-q+1$$

and the same result could have been obtained by pooling the degrees of freedom associated with SS(TB) and SSE.

An examination of the derivation of the expected values of MST and MSB in section 4.4 shows, that to obtain the expected values for the present case, all we need do is set all terms involving the $(\tau\beta)_{ij}$'s equal to zero. We have

$$E(SSE_1) = E(SSE) + E[SS(TB)]$$
$$= [(N-pq) + (p-1)(q-1)]\sigma^2 = (N-p-q+1) \quad ,$$

a result obtained by setting the $(\tau\beta)_{ij} = 0$ in $E[SS(TB)]$, and hence

$$E(MSE_1) = \sigma^2.$$

In the tests of section 4.1, for the Type I model, SSE_1 plays the role of SSE.

4.6 Distributions of the Sums of Squares

Corresponding to the hypotheses

$$H_1: (\tau\beta)_{ij} = 0 \, , \quad H_2: \tau_i = 0 \quad , \quad H_3: \beta_j = 0 \quad ,$$

we have the hypotheses

$$\sigma_{\tau\beta} = 0 \quad , \quad \sigma_\tau = 0 \quad , \quad \sigma_\beta = 0 \quad ,$$

if the corresponding variables are from other than a Type I population. Then, since the populations have zero means, it follows that the corresponding variables are equal to zero.

Returning to the Type I model, the theory of Chapter II implies that

$$t_i = \bar{Y}_{i..} - \bar{Y}_{...} \, , \quad b_j = \bar{Y}_{.j.} - \bar{Y}_{...} \, , \quad (tb)_{ij} = \bar{Y}_{ij.} - \bar{Y}_{.j.} - \bar{Y}_{i..} + \bar{Y}_{...} \, ,$$

$(i = 1,2,\ldots,p-1; j = 1,2,\ldots,q-1)$, are distributed independently of SSE. Therefore any function of these statistics is distributed independently of SSE and, in particular, this holds for t_p, b_q, $(tb)_{pj}$ $(j = 1,2,\ldots,q)$ and $(tb)_{iq}$ $(i = 1,2,\ldots,p-1)$. These results were obtained for the model

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk_{ij}}$$

and they hold for the particular case where $Y_{ijk_{ij}} = \varepsilon_{ijk_{ij}}$.
Hence

$$\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...} \,, \quad \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...} \,, \quad \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{...}$$

$$(i = 1,2,\ldots,p; j = 1,2,\ldots,q)$$

are distributed independently of SSE.

We shall now show that any variable of the above three types is independent of any variable of the other two types. Since they have normal distributions, it is sufficient to prove that

$$\mathrm{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}, \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...}) = \mathrm{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}, \bar{\varepsilon}_{Ij.} - \bar{\varepsilon}_{I..} - \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{...})$$
$$= \mathrm{cov}(\bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...}, \bar{\varepsilon}_{iJ.} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.J.} + \bar{\varepsilon}_{...}) = 0 \,.$$

We first compute

$$E(\bar{\varepsilon}_{i..}\bar{\varepsilon}_{.j.}) = \frac{1}{n_{i.}n_{.j}}(\sigma^2 \times \text{the number of } \varepsilon_{ijk_{ij}}\text{'s that } \bar{\varepsilon}_{i..} \text{ and } \bar{\varepsilon}_{.j.} \text{ have in common})$$

$$= \frac{n_{ij}}{n_{i.}n_{.j}}\sigma^2 = \frac{\sigma^2}{N} \,,$$

$$E(\bar{\varepsilon}_{i..}\bar{\varepsilon}_{...}) = \frac{1}{n_{i.}N} n_{i.}\sigma^2 = \frac{\sigma^2}{N} \,,$$

$$E(\bar{\varepsilon}_{.j.}\bar{\varepsilon}_{...}) = \frac{1}{n_{.j}N} n_{.j}\sigma^2 = \frac{\sigma^2}{N} \,,$$

$$E(\bar{\varepsilon}^2_{...}) = \frac{\sigma^2}{N} \,,$$

and

$$\mathrm{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}, \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...}) = \sigma^2\left(\frac{1}{N} - \frac{1}{N} - \frac{1}{N} + \frac{1}{N}\right) = 0 \,.$$

Next,

$$\operatorname{cov}\left(\bar{\varepsilon}_{i..}-\bar{\varepsilon}_{...},\ \bar{\varepsilon}_{Ij.}-\bar{\varepsilon}_{I..}-\bar{\varepsilon}_{.j.}+\bar{\varepsilon}_{...}\right)=\delta_{iI}\left(\frac{1}{n_{i.}n_{ij}}n_{ij}\sigma^2-\frac{\sigma^2}{n_{i.}}\right)$$

$$+\ \sigma^2\left(-\frac{1}{N}+\frac{1}{N}-\frac{1}{n_{Ij}N}n_{Ij}+\frac{1}{N}+\frac{1}{N}-\frac{1}{N}\right)=0\ ,$$

and similarly

$$\operatorname{cov}\left(\bar{\varepsilon}_{.j.}-\bar{\varepsilon}_{...},\ \bar{\varepsilon}_{iJ.}-\bar{\varepsilon}_{i..}-\bar{\varepsilon}_{.J.}+\bar{\varepsilon}_{...}\right)=0\ .$$

For the Type I model with interaction we saw in section 4.1 that the appropriate tests for H_1, H_2, and H_3 were

$$F_1=\frac{MS(TB)}{MSE}\ ,\quad F_2=\frac{MST}{MSE}\ ,\quad F_3=\frac{MSB}{MSE}$$

since, subject to the corresponding hypotheses,

$$SS(TB)=\sum_{i,j}^{p,q}n_{ij}\left(\bar{\varepsilon}_{ij.}-\bar{\varepsilon}_{i..}-\bar{\varepsilon}_{.j.}+\bar{\varepsilon}_{...}\right)^2,\quad SST=\sum_{i=1}^{p}n_{i.}\left(\bar{\varepsilon}_{i..}-\bar{\varepsilon}_{...}\right)^2,$$

$$SSB=\sum_{j=1}^{q}n_{.j}\left(\bar{\varepsilon}_{.j.}-\bar{\varepsilon}_{...}\right)^2\ ,\quad SSE=\sum_{i,j,k_{ij}}^{p,q,n_{ij}}\left(\varepsilon_{ijk_{ij}}-\bar{\varepsilon}_{ij.}\right)^2$$

are independently distributed as $\chi^2\sigma^2$ with (p-1)(q-1), p-1, q-1, and N-pq d.f. , respectively.

For the Type I model with no interaction we replace MSE by MSE_1 since then SST, SSB, and SSE_1 are independently distributed as $\chi^2\sigma^2$ with p-1,q-1, and N-p-q + 1 d.f. , respectively.

Our problem is to determine what tests can be made when we are not dealing with a Type I model. We recall that

$$SST=\sum_{i=1}^{p}n_{i.}\left[\tau_i-\bar{\tau}_.+(\overline{\tau\beta})_{i.}-(\overline{\tau\beta})_{..}+\bar{\varepsilon}_{i..}-\bar{\varepsilon}_{...}\right]^2,$$

$$SSB = \sum_{j=1}^{g} n_{\cdot j} \left[\beta_j - \bar{\beta}_{\cdot} + (\overline{\tau\beta})_{\cdot j} - (\overline{\tau\beta})_{\cdot\cdot} + \bar{\varepsilon}_{\cdot j \cdot} - \bar{\varepsilon}_{\cdot\cdot\cdot} \right]^2$$

and

$$SS(TB) = \sum_{i,j}^{p,g} n_{ij} \left[(\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} - (\overline{\tau\beta})_{\cdot j} + (\overline{\tau\beta})_{\cdot\cdot} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot j \cdot} + \bar{\varepsilon}_{\cdot\cdot\cdot} \right]^2 .$$

If there is no interaction term, subject to H_2, H_3, the sums of squares SST and SSB reduce to the corresponding expressions for the Type I model and the same tests apply. If there is an interaction term, the above argument shows that the Type I test can be used for H_1. Thus our problem is reduced to testing H_2 and H_3 when there is interaction and we are not dealing with a Type I model.

We first consider the Type II model where the τ_i's, β_j's, $(\tau\beta)_{ij}$'s are NID with zero means and variances σ_τ^2, σ_β^2, $\sigma_{\tau\beta}^2$, respectively. When H_2: $\sigma_\tau = 0$ holds, let

$$y_i = (\overline{\tau\beta})_{i\cdot} + \bar{\varepsilon}_{i\cdot\cdot} = \frac{1}{N} \sum_{j=1}^{g} n_{ij} (\tau\beta)_{ij} + \bar{\varepsilon}_{i\cdot\cdot} .$$

Then

$$\bar{y}_{\cdot} = \frac{1}{N} \sum_{i=1}^{p} n_{i\cdot} y_i = \frac{1}{N^2} \sum_{i,j}^{p,g} n_{i\cdot} n_{ij} (\tau\beta)_{ij} + \frac{1}{N} \sum_{i=1}^{p} n_{i\cdot} \bar{\varepsilon}_{i\cdot\cdot}$$

$$= \frac{1}{N} \sum_{i,j}^{p,g} n_{ij} (\tau\beta)_{ij} + \bar{\varepsilon}_{\cdot\cdot\cdot} = (\overline{\tau\beta})_{\cdot\cdot} + \bar{\varepsilon}_{\cdot\cdot\cdot} ,$$

and

$$SST = \sum_{i=1}^{p} n_{i\cdot} (y_i - \bar{y}_{\cdot})^2$$

where

$$E(y_i)=0\ ,\quad Var(y_i)=\frac{\sigma_{\tau\beta}^2}{N^2}\sum_{j=1}^{q} n_{ij}^2+\frac{\sigma^2}{n_{i.}}\ ,\quad cov(y_i,y_{i'})=0\ .$$

In Chapter III we had a similar situation and found that

SST/E(MST)

did not, in general, have a χ^2 distribution unless we let $n_i = n$. Accordingly, we begin again, assuming $n_{ij} = n$ and find that

$$n_{i.} = qn\ ,\quad n_{.j} = pn\ ,\quad N = pqn\ ,$$

$$\bar{\tau}_{.} = \frac{1}{N}\sum_{i=1}^{p} n_{i.}\,\tau_i = \frac{1}{p}\sum_{i=1}^{p}\tau_i\ ,$$

$$(\overline{\tau\beta})_{i.} = \frac{1}{N}\sum_{j=1}^{q} n_{.j}(\tau\beta)_{ij} = \frac{1}{q}\sum_{j=1}^{q}(\tau\beta)_{ij}\ ,\quad (\overline{\tau\beta})_{..} = \frac{1}{pq}\sum_{i,j}^{p,q}(\tau\beta)_{ij}\ .$$

We no longer impose the restriction that H_2 holds and let

$$y_i = \tau_i + (\overline{\tau\beta})_{i.} + \bar{\varepsilon}_{i..}\ ,$$

$$\bar{y}_{.} = \frac{1}{p}\sum_{i=1}^{p}\{\tau_i + (\overline{\tau\beta})_{i.} + \bar{\varepsilon}_{i..}\} = \bar{\tau}_{.} + (\overline{\tau\beta})_{..} + \bar{\varepsilon}_{...}\ .$$

Then

$$E(y_i)=0\ ,\quad Var(y_i) = \sigma_\tau^2 + \frac{\sigma_{\tau\beta}^2}{q} + \frac{\sigma^2}{qn}\ ,\quad cov(y_i,y_{i'})=0\ ,$$

$$SST = \sum_{i=1}^{p} qn\,(y_i-\bar{y}_{.})^2\ ,$$

and, by section 4.4,

$$E(MST) = \sigma^2 + n\,\sigma_{\tau\beta}^2 + qn\,\sigma_\tau^2\ .$$

Hence

$$\frac{SST}{E(MST)} = \frac{\sum_{i=1}^{p}(y_i - \bar{y})^2}{\sigma_\tau^2 + \frac{\sigma_{\tau\beta}^2}{q} + \frac{\sigma^2}{qn}} ,$$

by Chapter II, has a χ^2 distribution with p-1 d.f. . Similarly, SSB is distributed as χ^2E(MSB) with q-1 d.f. . Consider the three sets of variables

$$(\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{..}\,,\ (\overline{\tau\beta})_{.j} - (\overline{\tau\beta})_{..}\,,\ (\tau\beta)_{ij} - (\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{.j} + (\overline{\tau\beta})_{..} \quad .$$

We have

$$\mathrm{cov}\{(\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{..}\,,\ (\overline{\tau\beta})_{.j} - (\overline{\tau\beta})_{..}\} = E[(\overline{\tau\beta})_{i.}(\overline{\tau\beta})_{.j}] - E[(\overline{\tau\beta})_{i.}(\overline{\tau\beta})_{..}]$$

$$- E[(\overline{\tau\beta})_{.j}(\overline{\tau\beta})_{..}] + E[(\overline{\tau\beta})^2_{..}]$$

$$= \frac{1}{pq}\sum_{i',j'}^{p,q} E[(\tau\beta)_{ij'}(\tau\beta)_{i'j}] - \frac{1}{pq^2}\sum_{i',j,j'}^{p,q} E[(\tau\beta)_{ij}(\tau\beta)_{i'j'}]$$

$$-\frac{1}{p^2q}\sum_{i,i',j}^{p,q} E[(\tau\beta)_{ij}(\tau\beta)_{i'j'}] + \frac{1}{p^2q^2}\sum_{i,i',j,j'}^{p,q} E[(\tau\beta)_{ij}(\tau\beta)_{i'j'}]$$

$$= \frac{\sigma_{\tau\beta}^2}{pq} - q\frac{\sigma_{\tau\beta}^2}{pq^2} - \frac{p\,\sigma_{\tau\beta}^2}{p^2q} + \frac{pq\,\sigma_{\tau\beta}^2}{p^2q^2} = 0 ,$$

$$\mathrm{cov}\{[(\overline{\tau\beta})_{I.} - (\overline{\tau\beta})_{..}],[(\tau\beta)_{ij} - (\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{.j} + (\overline{\tau\beta})_{..}]\}$$

$$= \delta_{iI}\left\{\frac{\sigma_{\tau\beta}^2}{q} - q\frac{\sigma_{\tau\beta}^2}{q^2}\right\} - \frac{\sigma_{\tau\beta}^2}{pq} + \frac{q}{pq^2}\sigma_{\tau\beta}^2 - \frac{\sigma_{\tau\beta}^2}{pq} + \frac{q}{pq^2}\sigma_{\tau\beta}^2 + \frac{p}{p^2q}\sigma_{\tau\beta}^2 - \frac{pq}{p^2q^2}\sigma_{\tau\beta}^2 = 0,$$

and similarly

$$\mathrm{cov}\{[(\overline{\tau\beta})_{.J} - (\overline{\tau\beta})_{..}],[(\tau\beta)_{ij} - (\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{.j} + (\overline{\tau\beta})_{..}]\} = 0 ,$$

proving that the three sets of variables are independent. Bearing in mind that we proved the corresponding relations for the $\bar{\varepsilon}_{ij.}$'s earlier, it follows that the three sets of variables

$$\tau_i - \bar{\tau}_. + (\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{..} + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}\,, \quad \beta_j - \bar{\beta}_. + (\overline{\tau\beta})_{.j} - (\overline{\tau\beta})_{..} + \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...}\,,$$

$$(\tau\beta)_{ij} - (\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{.j} + (\overline{\tau\beta})_{..} + \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{...}\,,$$

are independently distributed and hence so are SST, SSB, and SS(TB).

If, in our Type I model, we set μ, the τ_i's and the β_j's equal to zero and assume the $(\tau\beta)_{ij}$'s are NID$(0, \sigma^2_{\tau\beta})$,

$$Y_{ijk} = (\tau\beta)_{ij} + \varepsilon_{ijk}\,, \qquad \bar{Y}_{ij.} = (\tau\beta)_{ij} + \bar{\varepsilon}_{ij.}\,,$$

$$E(\bar{Y}_{ij.}) = 0\,, \qquad Var(\bar{Y}_{ij.}) = \sigma^2_{\tau\beta} + \frac{\sigma^2}{n}$$

and the $\bar{Y}_{ij.}$'s are independent. We may carry out an analysis of variance on the $\bar{Y}_{ij.}$'s. According to the model of section 4.5, where there is no interaction, with the N of that section equal to pq and $n_{ij} = 1$,

$$\begin{aligned} SSE_1 &= SSE + SS(TB) \\ &= \sum_{i,j,k}^{p,q,1} (Y_{ijk} - \bar{Y}_{ij.})^2 + \sum_{i,j}^{p,q} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \\ &= \sum_{i,j}^{p,q} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \end{aligned}$$

since, under these conditions, $Y_{ijk} = \bar{Y}_{ij.}$. The theory of section 4.5 tells us that SSE_1 is distributed as $\chi^2(\sigma^2_{\tau\beta} + \frac{\sigma^2}{n})$ and hence

$$SS(TB) = \sum_{i,j}^{p,q} n(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

is distributed as $\chi^2(n\sigma^2_{\tau\beta} + \sigma^2)$ with pq-p-q +1 = (p-1)(q-1) degrees of freedom.

We display the above results in table 4.3 .

TABLE 4.3

Source of Variation	Degrees Of Freedom	Sum of Squares	Mean Square	E(MS)
Treatments	p-1	SST	MST	$\sigma^2 + n\sigma_{\tau\beta}^2 + qn\sigma_\tau^2$
Blocks	q-1	SSB	MSB	$\sigma^2 + n\sigma_{\tau\beta}^2 + pn\sigma_\beta^2$
Interaction	(p-1)(q-1)	SS(TB)	MS(TB)	$\sigma^2 + n\sigma_{\tau\beta}^2$
Error	N-pq	SSE	MSE	σ^2
Total	N-1			

We can determine the appropriate tests for H_1, H_2, and H_3, by examining the expected mean squares column. We know that the sums of squares, divided by the expected values of their mean squares are independently distributed as χ^2. If we also divide by the corresponding degrees of freedom, the ratio of any two has the F distribution. However, the computation can be carried out only when the expected mean squares cancel out in this ratio. Thus we have

$$F_1 = \frac{MS(TB)}{MSE}, \quad F_2 = \frac{MST}{MS(TB)}, \quad F_3 = \frac{MSB}{MS(TB)}$$

to test the hypothese H_1, H_2, H_3, respectively. All of these results are for the case where $n_{ij} = n$. If this condition does not hold, we can carry out the test for H_1 only. It will be noted that the above tests for H_2 and H_3 are different from the corresponding tests for the Type I model where MSE is the denominator used.

For a Type III model with interaction we can not expect to obtain the χ^2 distributions necessary for F tests of H_2 and H_3, in fact, any such test must depend on the $(\tau\beta)_{ij}$'s being normally distributed, whatever the nature of the model, unless the terms involving the $(\tau\beta)_{ij}$'s reduce to zero as in the Type I model.

An approach similar to the one given above would be used in the case of a mixed model.

CHAPTER V

TWO-WAY NESTED CLASSIFICATION MODELS

5.1 The Type I Model for Proportional Frequencies

In Chapter I we discussed an experiment in which the yields of naphthalene black for different samples of H acid were measured. It was assumed that the samples of H acid were random samples produced from naphthalene by a particular tar distiller. We shall assume that, in general, the experiment is carried out p times, the supplier of naphthalene being changed for each experiment. To describe the data we consider the model

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk_{ij}} , \tag{5.11}$$
$$(i = 1,2,\ldots,p; j = 1,2,\ldots,q; k_{ij} = 1,2,\ldots,n_{ij})$$

where the τ_i's represent the effects upon the yields associated with the various suppliers and the $\beta_{j(i)}$'s represent the effects due to variations between q samples from each of the suppliers. The $\varepsilon_{ijk_{ij}}$'s are NID($0,\sigma^2$). Associated with the jth sample of the ith supplier are n_{ij} observations and since we are considering proportional frequencies, we have

$$n_{ij} = \frac{n_{i.}n_{.j}}{N} .$$

The parameters are subject to the conditions

$$\sum_{i,j}^{p,q} n_{ij}\,\tau_i = \sum_{i=1}^{p} n_{i.}\,\tau_i = 0 , \tag{5.12}$$

$$\sum_{j=1}^{q} n_{ij}\,\beta_{j(i)} = \sum_{j=1}^{q} n_{.j}\,\beta_{j(i)} = 0 . \tag{5.13}$$

To show that these conditions can be satisfied, denote by ξ_{ij} the mean of a given subclass or sample, where

$$\xi_{ij} = \mu + \tau_i + \beta_{j(i)} \quad .$$

Then, if the conditions hold,

$$\bar{\xi}_{i.} = \frac{1}{n_{i.}} \sum_{j=1}^{q} n_{ij}\, \xi_{ij} = \mu + \tau_i \quad ,$$

$$\bar{\xi}_{..} = \frac{1}{N} \sum_{i,j}^{p,q} n_{ij}\, \xi_{ij} = \frac{1}{N} \sum_{i=1}^{p} n_{i.}\, \bar{\xi}_{i.} = \mu \quad .$$

Thus we must define

$$\mu = \bar{\xi}_{..} \quad , \quad \tau_i = \bar{\xi}_{i.} - \bar{\xi}_{..} \quad , \quad \beta_{j(i)} = \xi_{ij} - \bar{\xi}_{i.} \quad .$$

Defining the parameters in this way it may be verified that conditions (5.12) and (5.13) are satisfied. For example,

$$\sum_{i=1}^{p} n_{i.}\, \tau_i = \sum_{i=1}^{p} n_{i.}\, \bar{\xi}_{i.} - \bar{\xi}_{..} \sum_{i=1}^{p} n_{i.} = N\bar{\xi}_{..} - N\bar{\xi}_{..} = 0 \quad ,$$

and

$$\sum_{j=1}^{q} n_{ij}\, \beta_{j(i)} = \sum_{j=1}^{q} n_{ij}\, \xi_{ij} - n_{i.}\, \bar{\xi}_{i.} = n_{i.}\, \bar{\xi}_{i.} - n_{i.}\, \bar{\xi}_{i.} = 0 \quad .$$

The equations (5.11) may be put in the form

$$(5.14) \qquad Y_{ijk_{ij}} = \mu + \sum_{i'=1}^{p} U_{i'i}\, \tau_{i'} + \sum_{i',j'}^{p,q} V_{j'(i')j(i)}\, \beta_{j'(i')} + \varepsilon_{ijk_{ij}}$$

where

$$U_{i'i} = \delta_{i'i} \quad , \quad V_{j'(i')j(i)} = \delta_{ii'}\, \delta_{jj'} \quad .$$

If we order the $Y_{ijk_{ij}}$'s in some way, calling them Y_α ($\alpha = 1, 2, \ldots, N$), the equations (5.14) may be put in the vector form

$$(5.15) \qquad Y = \mu + \sum_{i'=1}^{p} U_{i'}\, \tau_{i'} + \sum_{i',j'}^{p,q} V_{j'(i')}\, \beta_{j'(i')} + \varepsilon$$

where the $U_{i'}$'s and $V_{j'(i')}$'s are the coefficient vectors of the $\tau_{i'}$'s and $\beta_{j'(i')}$'s, respectively.

Denoting the elements of $U_{i'}$ by $U_{i'\alpha}$ $(\alpha = 1,2,\ldots,N)$, we define

$$\bar{U}_{i'} = \frac{1}{N}\sum_{\alpha=1}^{N} U_{i'\alpha} = \frac{1}{N}\sum_{i,j,k_{ij}}^{p,g,n_{ij}} U_{i'i} = \frac{1}{N}\sum_{i,j}^{p,g} n_{ij}\,\delta_{i'i} = \frac{n_{i'.}}{N} ,$$

and $u_{i'\alpha} = U_{i'\alpha} - \bar{U}_{i'}$, so that

$$0 = \sum_{\alpha=1}^{N} u_{i'\alpha} = \sum_{i,j,k_{ij}}^{p,g,n_{ij}} u_{i'i} = \sum_{i,j}^{p,g} n_{ij}\,u_{i'i} = \sum_{i=1}^{p} n_{i.}\,u_{i'i} .$$

Define

$$\bar{V}_{j'(i')} = \frac{1}{N}\sum_{\alpha=1}^{N} V_{j'(i')\alpha} = \frac{1}{N}\sum_{i,j,k_{ij}}^{p,g,n_{ij}} V_{j'(i')j(i)} = \frac{1}{N}\sum_{i,j}^{p,g} n_{ij}\,\delta_{ii'}\,\delta_{jj'} = \frac{n_{i'j'}}{N} ,$$

and $v_{j'(i')\alpha} = V_{j'(i')\alpha} - \bar{V}_{j'(i')}$.

Then

$$\sum_{\alpha=1}^{N} v_{j'(i')\alpha} = \sum_{i,j}^{p,g} n_{ij}\,v_{j'(i')j(i)} = 0 .$$

Now denote by $\bar{U}_{i'}$ and $\bar{V}_{j'(i')}$ the vectors $\bar{U}_{i'}I$ and $\bar{V}_{j'(i')}I$ respectively, and set

$$u_{i'} = U_{i'} - \bar{U}_{i'} , \quad v_{j'(i')} = V_{j'(i')} - \bar{V}_{j'(i')} .$$

We may then write (5.15) in the form

$$(5.16) \qquad Y = \mu + \sum_{i'=1}^{p} (u_{i'} + \bar{U}_{i'})\,\tau_{i'} + \sum_{i',j'}^{p,g} (v_{j'(i')} + \bar{V}_{j'(i')})\,\beta_{j'(i')} + \varepsilon$$

$$= \mu + \sum_{i'=1}^{p} u_{i'}\,\tau_{i'} + \sum_{i',j'}^{p,g} v_{j'(i')}\,\beta_{j'(i')} + \varepsilon$$

since

$$\frac{1}{N}\sum_{i'=1}^{p} n_{i'.}\,\tau_{i'} = \frac{1}{N}\sum_{j'=1}^{g} n_{i'j'}\,\beta_{j'(i')} = 0 .$$

To apply the theory of Chapter II, it is necessary that the $u_{i'}$'s and $v_{j'}$'s form a linearly independent set of vectors. We shall show that this is not the case. Using the methods of Chapter IV we shall be able to remedy this situation. First, however we need certain relations among the vectors. We now show that

$$\sum_{i'=1}^{p} U_{i'} = I$$

This follows since, when we add the elements in row α of these vectors, we have

$$\sum_{i'=1}^{p} U_{i'\alpha} = \sum_{i'=1}^{p} U_{i'i} = \sum_{i'=1}^{p} \delta_{ii'} = 1 \quad .$$

Similarly we can show that

$$\sum_{i',j'}^{p,g} V_{j'(i')} = I \quad .$$

Also

$$U_i \cdot I = n_{i.} \quad \text{and} \quad V_{j(i)} \cdot I = n_{ij}$$

and the following multiplication table gives us the values of different dot products formed from our vectors:

	U_i	V
$U_{i'}$	$n_{i.}\,\delta_{ii'}$	$n_{ij}\,\delta_{ii'}$
$V_{j'(i')}$	$n_{ij'}\,\delta_{ii'}$	$n_{ij}\,\delta_{ii'}\,\delta_{jj'}$

Now

$$\sum_{i=1}^{p} u_i = \sum_{i=1}^{p} U_i - \sum_{i=1}^{p} \overline{U}_i = 0$$

since

$$\sum_{i=1}^{p} \bar{U}_i = I \sum_{i=1}^{p} \frac{n_{i\cdot}}{N} = I \ .$$

Hence the u_i's and $v_{j(i)}$'s do not form a linearly independent set. To meet this difficulty we use the relations (5.12) and (5.13) to eliminate τ_p and $\beta_{g(i)}$ (i = 1,2,...,p). We have

$$\tau_p = -\frac{1}{n_{p\cdot}} \sum_{i=1}^{p-1} n_{i\cdot}\,\tau_i \quad ,$$

$$n_{ig}\,\beta_{g(i)} = -\sum_{j=1}^{g-1} n_{ij}\,\beta_{j(i)} \qquad (i = 1, 2, \cdots, p) \ .$$

Therefore

$$\sum_{i'=1}^{p} u_{i'}\tau_{i'} = \sum_{i'=1}^{p-1} u_{i'}\tau_{i'} - \frac{u_p}{n_{p\cdot}} \sum_{i'=1}^{p-1} n_{i'\cdot}\,\tau_{i'} = \sum_{i'=1}^{p-1} \left(u_{i'} - \frac{n_{i'\cdot}}{n_{p\cdot}} u_p \right) \tau_{i'} \quad ,$$

$$\sum_{j'=1}^{g} v_{j'(i')}\,\beta_{j'(i')} = \sum_{j'=1}^{g-1} v_{j'(i')}\,\beta_{j'(i')} - \frac{v_{g(i')}}{n_{i'g}} \sum_{j'=1}^{g-1} n_{i'j'}\,\beta_{j'(i')}$$

$$= \sum_{j'=1}^{g-1} \left(v_{j'(i')} - \frac{n_{i'j'}}{n_{i'g}} v_{g(i')} \right) \beta_{j'(i')} \quad .$$

We can now write equation (5.16) in the form

$$(5.17) \qquad Y = \mu + \sum_{i'=1}^{p-1} \left(u_{i'} - \frac{n_{i'\cdot}}{n_{p\cdot}} u_p \right) \tau_{i'} + \sum_{i'=1}^{p} \sum_{j'=1}^{g-1} \left(v_{j'(i')} - \frac{n_{i'j'}}{n_{i'g}} v_{g(i')} \right) \beta_{j'(i')} + \varepsilon$$

Note that

$$u_{i'} - \frac{n_{i'\cdot}}{n_{p\cdot}} u_p = U_{i'} - \bar{U}_{i'} - \frac{n_{i'\cdot}}{n_{p\cdot}} (U_p - \bar{U}_p)$$

$$= U_{i'} - \frac{n_{i'\cdot}}{n_{p\cdot}} U_p - \frac{n_{i'\cdot}}{N} I + \frac{n_{i'\cdot}}{n_{p\cdot}} \frac{n_{p\cdot}}{N} I = U_{i'} - \frac{n_{i'\cdot}}{n_{p\cdot}} U_p \quad .$$

Also

$$v_{j'(i')} - \frac{n_{i'j'}}{n_{i'q}} v_{q(i')} = V_{j'(i')} - \bar{V}_{j'(i')} - \frac{n_{i'j'}}{n_{i'q}} \left(V_{q(i')} - \bar{V}_{q(i')} \right)$$

$$= V_{j'(i')} - \frac{n_{i'j'}}{n_{i'q}} V_{q(i')} - \frac{n_{i'j'}}{N} I + \frac{n_{i'j'}}{n_{i'q}} \frac{n_{i'q}}{N} I$$

$$= V_{j'(i')} - \frac{n_{i'j'}}{n_{i'q}} V_{q(i')} .$$

Consider the equation

$$\sum_{i'=1}^{p-1} c_{i'} \left(U_{i'} - \frac{n_{i'.}}{n_{p.}} U_p \right) = 0$$

Multiplying the equation by U_i $(i = 1,2,\ldots,p-1)$ we find that $n_{i.} c_i = 0$, since the U_i's are orthogonal vectors. Thus the vectors

$$U_{i'} - \frac{n_{i'.}}{n_{p.}} U_p \qquad (i' = 1, 2, \cdots, p-1)$$

are linearly independent. Next consider the equation

$$\sum_{i'=1}^{p} \sum_{j'=1}^{q-1} c_{j'(i')} \left(V_{j'(i')} - \frac{n_{i'j'}}{n_{i'q}} V_{q(i')} \right) = 0 .$$

Multiplying by $V_{j(i)}$ $(i = 1,2,\ldots,p; j = 1,2,\ldots,q-1)$ we find that $n_{ij} c_j = 0$ since the $V_{j'(i')}$'s are orthogonal. Thus the vectors

$$V_{j'(i')} - \frac{n_{i'j'}}{n_{i'q}} V_{q(i')} \qquad (j' = 1, 2, \cdots, q-1 ; i' = 1, 2, \cdots, p)$$

are linearly independent. Also

$$U_i \cdot \left(V_{j'(i')} - \frac{n_{i'j'}}{n_{i'q}} V_{q(i')} \right) = \delta_{ii'} n_{i'j'} - \delta_{ii'} \frac{n_{i'j'}}{n_{i'q}} n_{i'q} = 0 .$$

Therefore the two sets of vectors are orthogonal to each other. Our model (5.17) now satisfies the conditions required in Chapter II.

We shall be interested in testing the two hypotheses

H_1: $\beta_{j(i)} = 0 \quad (j = 1,2,\ldots,q-1; i = 1,2,\ldots,p)$,

H_2: $\tau_i = 0 \quad (i = 1,2,\ldots,p-1)$.

Conditions (5.12) and (5.13) together with these two hypotheses imply that all parameters of the kind appearing in the two hypotheses are zero.

To test H_1, we first compute

$$SSE = \sum_{i,j,k_{ij}}^{p,q,k_{ij}} [Y_{ijk_{ij}} - m - t_i - b_{j(i)}]^2 ,$$

where m, t_i and $b_{j(i)}$ are the least squares estimates of μ, τ_i and $\beta_{j(i)}$ respectively and SSE is the minimized value of the residual sum of squares. Next, we compute SSE_1, the corresponding minimum obtained under the assumption that H_1 holds. Then

$$R = \sum_{\alpha=1}^{N} y_\alpha^2 - SSE$$

is the reduction in the sum of squares when all the parameters are used. Also

$$R_1 = \sum_{\alpha=1}^{N} y_\alpha^2 - SSE_1$$

is the reduction due to the parameters left when H_1 is true. As in the preceding chapter, $R \geq R_1$ and the additional reduction in the sum of squares due to the $\beta_{j(i)}$'s is

$$SSB = R-R_1 = SSE_1 - SSE = \sum_{i,j}^{p,q-1} [b'_{j(i)}]^2$$

where the $b'_{j(i)}$'s are the estimates obtained when the orthogonal model of Chapter II is used.

Similarly, SSE_2 denotes the minimum obtained subject to H_2 and the reduction in the sum of squares due to the τ_i's is

$$SST = R-R_2 = SSE_2 - SSE = \sum_{i=1}^{p-1} [t'_i]^2 .$$

Finally,

$$SSE = \sum_{\alpha=1}^{N} y_\alpha^2 - R$$

$$= \sum_{\alpha=1}^{N} y_\alpha^2 - \sum_{i=1}^{p-1} [t'_i]^2 - \sum_{i,j}^{p,q-1} [b'_{j(i)}]^2 ,$$

so that

$$\sum_{\alpha=1}^{N} y_\alpha^2 = SST + SSB + SSE .$$

From Chapter II we know that SST, SSB, and SSE are independently distributed as $\chi^2\sigma^2$ with p-1, p(q-1) and N-pq degrees of freedom, respectively, under the corresponding hypotheses. The hypotheses H_1 and H_2 are tested by the statistics

$$F_1 = \frac{MSB}{MSE} \quad \text{and} \quad F_2 = \frac{MST}{MSE} ,$$

respectively.

In the next section we shall develop methods for the computation of these statistics.

5.2 The Sums of Squares

Our estimates of μ, τ_i, $\beta_{j(i)}$ are m, t_i and $b_{j(i)}$ respectively, where these values minimize

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - m - t_i - b_{j(i)})^2$$

subject to the conditions

(5.21) $$\sum_{i=1}^{p} n_{ij} t_i = \sum_{i=1}^{p} n_{i.} t_i = 0 \quad ,$$

(5.22) $$\sum_{j=1}^{q} n_{ij} b_{j(i)} = \sum_{j=1}^{q} n_{ij} b_{j(i)} = 0 \qquad (i = 1, 2, \cdots, p) \; .$$

By Theorem 3.1, we can ignore condition (5.21) since its Lagrange multiplier will be zero. However condition (5.22) will have to be considered in the computation of SSE_2. We shall compute SSE in the same manner as in Chapter IV.

We have

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \xi_{ij})^2$$

where $E(Y_{ijk_{ij}}) = \xi_{ij}$. Then

$$\frac{\partial SSE}{\partial \xi_{ij}} = -2 \sum_{k_{ij}=1}^{n_{ij}} (Y_{ijk_{ij}} - \xi_{ij}) = 0 \quad ,$$

and our estimate of ξ_{ij} is $\hat{\xi}_{ij} = \overline{Y}_{ij.}$. Then, by the invariance property of such estimators,

$$m = \bar{\hat{\xi}}_{..} = \frac{1}{N} \sum_{i,j}^{p,q} \hat{\xi}_{ij} = \frac{1}{N} \sum_{i,j,k_{ij}}^{p,q,n_{ij}} Y_{ijk_{ij}} = \overline{Y}_{...} \quad ,$$

$$t_i = \bar{\hat{\xi}}_{i.} - m = \frac{1}{n_{i.}} \sum_{j=1}^{q} n_{ij} \overline{Y}_{ij.} - m = \overline{Y}_{i..} - \overline{Y}_{...} \quad ,$$

$$b_{j(i)} = \hat{\xi}_{ij} - \bar{\hat{\xi}}_{i.} = \bar{Y}_{ij.} - \bar{Y}_{i..} \quad ,$$

and

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})^2 \quad .$$

To obtain SSE_1, we must minimize

$$SSE_1 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - m - t_i)^2 \quad .$$

We have

$$\frac{\partial SSE_1}{\partial m} = -2 \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - m - t_i) = 0 \quad ,$$

$$\frac{\partial SSE_1}{\partial t_i} = -2 \sum_{j,k_{ij}}^{q,n_{ij}} (Y_{ijk_{ij}} - m - t_i) = 0 \quad ,$$

and making use of condition (5.21) we find

$$m = \bar{Y}_{...} \quad \text{and} \quad t_i = \bar{Y}_{i..} - \bar{Y}_{...} \quad .$$

Therefore

$$SSE_1 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{i..})^2$$

Then

$$
\begin{aligned}
SSB &= SSE_1 - SSE \\
&= \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{i..})^2 - \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})^2 \\
&= \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (\bar{Y}_{ij.} - \bar{Y}_{i..})^2 + 2 \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})(\bar{Y}_{ij.} - \bar{Y}_{i..}) \\
&= \sum_{i,j}^{p,q} n_{ij} (\bar{Y}_{ij.} - \bar{Y}_{i..})^2
\end{aligned}
$$

since the second sum is equal to

$$2\sum_{ij}^{p,q} n_{ij}(\bar{Y}_{ij.} - \bar{Y}_{ij.})(\bar{Y}_{ij.} - \bar{Y}_{i..}) = 0 \quad .$$

To determine SSE_2, we minimize

$$SSE_2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - m - b_{j(i)})^2$$

subject to condition (5.22). Thus we must minimize the expression

$$Q = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - m - b_{j(i)})^2 + \sum_{i=1}^{p} c_i \sum_{j=1}^{q} n_{ij}\, b_{j(i)} \quad .$$

Taking partial derivatives with respect to m and b_j we have

$$(5.23) \qquad \frac{\partial Q}{\partial m} = -2\sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - m - b_{j(i)}) = 0$$

$$(5.24) \qquad \frac{\partial Q}{\partial b_{j(i)}} = -2\sum_{k_{ij}}^{n_{ij}} (Y_{ijk_{ij}} - m - b_{j(i)}) + c_i n_{ij} = 0 \quad .$$

From (5.23) we find $m = \bar{Y}_{...}$ and from (5.24)

$$(5.25) \qquad \bar{Y}_{ij.} - m - b_{j(i)} - \frac{c_i N}{2n_{i.}} = 0 \quad .$$

Multiplying by n_{ij} and summing over j we have

$$Y_{i..} - n_{i.}m - \frac{c_i N}{2} = 0$$

and hence

$$c_i = \frac{2n_{i.}}{N}(\bar{Y}_{i..} - \bar{Y}_{...}) \quad .$$

Substituting for c_i in (5.25) we obtain

$$\bar{Y}_{ij.} - \bar{Y}_{...} - b_{j(i)} - (\bar{Y}_{i..} - \bar{Y}_{...}) = 0 \quad .$$

Therefore

$$b_{j(i)} = \bar{Y}_{ij.} - \bar{Y}_{i..}$$

and

$$SSE_2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.} + \bar{Y}_{i..} - \bar{Y}_{...})^2 \quad .$$

Then

$$SST = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.} + \bar{Y}_{i..} - \bar{Y}_{...})^2 - \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})^2$$

$$= \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (\bar{Y}_{i..} - \bar{Y}_{...})^2 + 2\sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})(\bar{Y}_{i..} - \bar{Y}_{...})$$

$$= \sum_{i=1}^{p} n_{i.} (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

since the second sum is equal to

$$2 \sum_{i,j}^{p,q} n_{ij} (\bar{Y}_{ij.} - \bar{Y}_{ij.})(\bar{Y}_{i..} - \bar{Y}_{...}) = 0 \quad .$$

5.3 Other Models

We still assume that

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk_{ij}} \quad .$$

For the Type II model we assume that the τ_i's, $\beta_{j(i)}$'s and $\varepsilon_{ijk_{ij}}$'s are NID with zero means and variances σ_τ^2, σ_β^2, and σ^2 respectively. We then have $E(Y_{ijk_{ij}}) = \mu$ and

$$Var(Y_{ijk_{ij}}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma^2 \quad .$$

The Type III model, as considered in Chapter IV, would not be realistic here. Corresponding to it, we have the case where the τ_i's come from a finite population of size P, mean zero, and variance

$$\sigma_\tau^2 = \frac{1}{P-1} \sum_{i=1}^{P} \tau_i^2 \quad ,$$

while the $\beta_{j(i)}$'s come from P populations, corresponding to the different values of i, these populations being independent of each other and the population of τ_i's, with zero means and common variance

$$\sigma_\beta^2 = \frac{1}{Q-1} \sum_{j=1}^{Q} \beta_{j(i)}^2 .$$

We shall still call this model the Type III model. The assumption of zero means implies that

$$\sum_{i=1}^{P} \tau_i = 0 \quad , \quad \sum_{j=1}^{Q} \beta_{j(i)} = 0 .$$

5.4 The Expected Values of the Sums of Squares

As before we shall arbitrarily begin with the sums of squares obtained for the Type I model and, since

$$\bar{Y}_{ij\cdot} = \mu + \tau_i + \beta_{j(i)} + \bar{\varepsilon}_{ij\cdot} ,$$

$$\bar{Y}_{i\cdot\cdot} = \mu + \tau_i + \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{i\cdot\cdot} ,$$

$$\bar{Y}_{\cdot\cdot\cdot} = \mu + \bar{\tau}_{\cdot} + \bar{\beta}_{\cdot(\cdot)} + \bar{\varepsilon}_{\cdot\cdot\cdot} ,$$

where

$$\bar{\tau}_{\cdot} = \frac{1}{N} \sum_{i=1}^{P} n_{i\cdot} \tau_i \quad , \quad \bar{\beta}_{\cdot(i)} = \frac{1}{N} \sum_{j=1}^{q} n_{\cdot j} \beta_{j(i)} ,$$

$$\bar{\beta}_{\cdot(\cdot)} = \frac{1}{N} \sum_{i,j}^{P,q} n_{ij} \beta_{j(i)} ,$$

$$SST = \sum_{i=1}^{P} n_{i\cdot} (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$= \sum_{i=1}^{P} n_{i\cdot} [\tau_i - \bar{\tau}_{\cdot} + \bar{\beta}_{\cdot(i)} - \bar{\beta}_{\cdot(\cdot)} + \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot\cdot\cdot}]^2 ,$$

$$SSB = \sum_{i,j}^{p,q} n_{ij} (\bar{Y}_{ij.} - \bar{Y}_{i..})^2$$

$$= \sum_{i,j}^{p,q} n_{ij} \left[\beta_{j(i)} - \bar{\beta}_{.(i)} + \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..} \right]^2 ,$$

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})^2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (\varepsilon_{ijk_{ij}} - \bar{\varepsilon}_{ij.})^2 .$$

To use Theorem 3.2 to evaluate the expected values of the above sums of squares we shall need the variances and covariances of the following sets of variables,

$$\tau_i + \bar{\beta}_{.(i)} + \bar{\varepsilon}_{i..} \quad , \quad \beta_{j(i)} + \bar{\varepsilon}_{ij.} \quad .$$

Hence, we must compute

$$E(\tau_i^2) , \; E(\tau_i \tau_{i'}) , \; E(\bar{\beta}_{.(i)}^2), \; E(\bar{\beta}_{.(i)} \bar{\beta}_{.(i')}) ,$$

$$E(\beta_{j(i)}^2) , \; E(\beta_{j(i)} \beta_{j'(i)}) , \; E(\beta_{j(i)} \beta_{j(i')}) ,$$

and $E(\beta_{j(i)} \beta_{j'(i')})$

in a form which will be valid regardless of the nature of the populations used.

If τ_i comes from a Type II population, $E(\tau_i^2) = \sigma_\tau^2$, and if it comes from a Type III population

$$E(\tau_i^2) = \frac{1}{p} \sum_{i=1}^{p} \tau_i^2 = \left(1 - \frac{1}{p}\right) \sigma_\tau^2 ,$$

which gives the formula for the Type II case if we let $P \to \infty$. For the Type II case, $E(\tau_i \tau_{i'}) = 0$, and for the Type III case

$$E(\tau_i \tau_{i'}) = \sum_{i \neq i'}^{P} \frac{\tau_i \tau_{i'}}{P(P-1)} = - \sum_{i=1}^{P} \frac{\tau_i^2}{P(P-1)} = - \frac{\sigma_\tau^2}{P}$$

since

$$0 = \sum_{i,i'}^{P} \tau_i \tau_{i'} = \sum_{i=1}^{P} \tau_i^2 + \sum_{i \neq i'}^{P} \tau_i \tau_{i'} .$$

Thus the formula for the Type II case is again included in the Type III case. Considering the $\beta_{j(i)}$'s for the Type III case,

$$E(\beta_{j(i)}^2) = \sum_{i,j}^{P,Q} \frac{\beta_{j(i)}^2}{PQ} = \frac{1}{P} \sum_{i=1}^{P} \frac{Q-1}{Q} \sigma_\beta^2 = \left(1 - \frac{1}{Q}\right) \sigma_\beta^2 ,$$

$$E(\beta_{j(i)} \beta_{j(i')}) = 0 ,$$

$$E(\beta_{j(i)} \beta_{j'(i)}) = \sum_{i,j \neq j'}^{P,Q} \frac{\beta_{j(i)} \beta_{j'(i)}}{PQ(Q-1)} = - \sum_{i,j}^{P,Q} \frac{\beta_{j(i)}^2}{PQ(Q-1)} = - \frac{\sigma_\beta^2}{Q} ,$$

since

$$0 = \sum_{j,j'}^{Q} \beta_{j(i)} \beta_{j'(i)} = \sum_{j=1}^{Q} \beta_{j(i)}^2 + \sum_{j \neq j'}^{Q} \beta_{j(i)} \beta_{j'(i)} ,$$

and $E(\beta_{j(i)} \beta_{j'(i')}) = 0$.

These results may be summarized in the single formula

$$E(\beta_{j(i)} \beta_{j'(i')}) = \delta_{ii'} \left(\delta_{jj'} - \frac{1}{Q}\right) \sigma_\beta^2 .$$

If we let $Q \rightarrow \infty$, this formula gives us the correct result for the Type II case since then

$$E(\beta_{j(i)}\beta_{j'(i')}) = \delta_{ii'}\delta_{jj'}\sigma_\beta^2 \ .$$

Next we consider the expected values of expressions involving $\bar{\beta}_{\cdot(i)}$. For the Type I model this term is zero, but for all other types of populations we find that

$$E(\bar{\beta}_{\cdot(i)}, \bar{\beta}_{\cdot(i')}) = 0 \ ,$$

$$E(\beta^2_{\cdot(i)}) = \frac{1}{N^2}\sum_{j,j'}^{g} n_{\cdot j}\, n_{\cdot j'}\, E(\beta_{j(i)}\beta_{j'(i)})$$

$$= \frac{1}{N^2}\sum_{j=1}^{g} n_{\cdot j}^2\, E(\beta^2_{j(i)}) + \frac{1}{N^2}\sum_{j\neq j'}^{g} n_{\cdot j}\, n_{\cdot j'}\, E(\beta_{j(i)}\beta_{j'(i)})$$

$$= \left(1-\frac{1}{Q}\right)\frac{1}{N^2}\sum_{j=1}^{g} n_{\cdot j}^2\,\sigma_\beta^2 - \frac{1}{N^2}\left(N^2-\sum_{j=1}^{g} n_{\cdot j}^2\right)\frac{\sigma_\beta^2}{Q}$$

$$= \frac{1}{N^2}\left[\sum_{j=1}^{g} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_\beta^2 \ ,$$

since

$$\sum_{j\neq j'}^{g} n_{\cdot j}\, n_{\cdot j'} = N^2 - \sum_{j=1}^{g} n_{\cdot j}^2 \ .$$

To evaluate E(SST), we let

$$y_i = \tau_i + \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{i\cdot\cdot} \ ,$$

$$\bar{y}_\cdot = \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}\, y_i = \bar{\tau}_\cdot + \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}\,\frac{1}{N}\sum_{j=1}^{g} n_{\cdot j}\,\beta_{j(i)} + \bar{\varepsilon}_{\cdots}$$

$$= \bar{\tau}_\cdot + \frac{1}{N}\sum_{i,j}^{p,g} n_{ij}\,\beta_{j(i)} + \bar{\varepsilon}_{\cdots} = \bar{\tau}_\cdot + \bar{\beta}_{\cdot(\cdot)} + \bar{\varepsilon}_{\cdots} \ ,$$

so that

$$SST = \sum_{i=1}^{p} n_{i\cdot}(y_i - \bar{y}_\cdot)^2 \ .$$

Now

$$\mu_i = E(y_{i\cdot}) = E(\tau_{i\cdot}) + \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}\, E(\beta_{j(i)})$$

$$= E(\tau_{i\cdot}) = (1-\delta_\tau)\,\tau_i \quad ,$$

$$\bar{\mu}_{\cdot} = \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}\,\mu_i = 0 \quad ,$$

where $\delta_\tau = 0$ if the τ_i's come from a Type I population and $\delta_\tau = 1$ otherwise. Also

$$\sigma_i^2 = Var(y_{i\cdot}) = E[\tau_i - E(\tau_i) + \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{i\cdot\cdot}]^2$$

$$= E[\tau_i - E(\tau_i)]^2 + E(\bar{\beta}^2_{\cdot(i)}) + E(\bar{\varepsilon}^2_{i\cdot\cdot})$$

$$= \delta_\tau\left(1-\frac{1}{P}\right)\sigma_\tau^2 + \delta_\beta\,\frac{1}{N^2}\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_\beta^2 + \frac{\sigma^2}{n_{i\cdot}} \quad ,$$

where $\delta_\beta = 0$ if the $\beta_{j(i)}$'s are from a Type I population and $\delta_\beta = 1$ otherwise. Next

$$\lambda = cov(y_{i\cdot}, y_{i'\cdot}) = E\{[\tau_i - E(\tau_i)][\tau_{i'} - E(\tau_{i'})]\} + E[\bar{\beta}_{\cdot(i)}\,\bar{\beta}_{\cdot(i')}]$$

$$= \delta_\tau\,\frac{\sigma_\tau^2}{P} \quad ,$$

$$\sigma_i^2 - \lambda = \delta_\tau\,\sigma_\tau^2 + \delta_\beta\,\frac{1}{N^2}\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_\beta^2 + \frac{\sigma^2}{n_{i\cdot}} \quad ,$$

$$E(SST) = (1-\delta_\tau)\sum_{i=1}^{p} n_{i\cdot}\,\tau_i^2 + \sum_{i=1}^{p} n_{i\cdot}\left(1-\frac{n_{i\cdot}}{N}\right)\Big\{\delta_\tau\,\sigma_\tau^2 +$$

$$\delta_\beta\,\frac{1}{N^2}\left[\sum_{j=1}^{8} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_\beta^2 + \frac{\sigma^2}{n_{i\cdot}}\Big\}$$

$$= (p-1)\sigma^2 + \delta_\beta \frac{1}{N^2}\left[N - \sum_{i=1}^{p} \frac{n_{i\cdot}^2}{N}\right]\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_\beta^2 +$$

$$\delta_\tau \left[N - \sum_{i=1}^{p} \frac{n_{i\cdot}^2}{N}\right]\sigma_\tau^2 + (1-\delta_\tau)\sum_{i=1}^{p} n_{i\cdot}\tau_i^2$$

and

$$E(MST) = \sigma^2 + \delta_\beta \frac{1}{(p-1)N}\left[N - \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}^2\right]\left[\sum_{j=1}^{q} n_{\cdot j}^2 - \frac{N^2}{Q}\right]\sigma_\beta^2 +$$

$$\delta_\tau \frac{1}{p-1}\left[N - \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}^2\right]\sigma_\tau^2 + \frac{(1-\delta_\tau)}{p-1}\sum_{i=1}^{p} n_{i\cdot}\tau_i^2 \ .$$

To evaluate E(SSB) we let

$$y_j = \beta_{j(i)} + \bar{\varepsilon}_{ij\cdot} \ , \quad \bar{y}_\cdot = \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j} y_j = \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{i\cdot\cdot} \ ,$$

so that

$$SSB = \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot} \sum_{j=1}^{q} n_{\cdot j}(y_j - \bar{y}_\cdot)^2 \ ,$$

$$\mu_j = E(y_j) = (1-\delta_\beta)\beta_{j(i)} \ , \quad \bar{\mu}_\cdot = \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}\mu_j = 0 \ ,$$

$$\begin{aligned}\sigma_j^2 = Var(y_j) &= E\{\beta_{j(i)} - (1-\delta_\beta)\beta_{j(i)} + \bar{\varepsilon}_{ij\cdot}\}^2 \\ &= E\{\delta_\beta \beta_{j(i)} + \bar{\varepsilon}_{ij\cdot}\}^2 \\ &= \delta_\beta E(\beta_{j(i)}^2) + E(\bar{\varepsilon}_{ij\cdot}^2) \\ &= \delta_\beta\left(1 - \frac{1}{Q}\right)\sigma_\beta^2 + \frac{\sigma^2}{n_{ij}} \ ,\end{aligned}$$

$$\lambda = \operatorname{cov}(y_j, y_{j'}) = E\{[\delta_\beta \beta_{j(i)} + \bar{\varepsilon}_{ij\cdot}][\delta_\beta \beta_{j'(i)} + \bar{\varepsilon}_{ij'\cdot}]\}$$

$$= \delta_\beta E(\beta_{j(i)} \beta_{j'(i)}) = -\delta_\beta \frac{\sigma_\beta^2}{Q} ,$$

$$\sigma_i^2 - \lambda = \delta_\beta \sigma_\beta^2 + \frac{\sigma^2}{n_{ij}} ,$$

$$E(SSB) = \frac{1}{N} \sum_{i=1}^{p} n_{i\cdot} \left\{ \sum_{j=1}^{q} n_{\cdot j}(1-\delta_\beta)\beta_{j(i)}^2 + \sum_{j=1}^{q} n_{\cdot j}\left(1 - \frac{n_{\cdot j}}{N}\right)\left(\delta_\beta \sigma_\beta^2 + \frac{\sigma^2}{n_{\cdot j}}\right) \right\}$$

$$= (1-\delta_\beta) \sum_{i,j}^{p,q} n_{ij} \beta_{j(i)}^2 + \delta_\beta \left[N - \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}^2\right] \sigma_\beta^2 + p(q-1)\sigma^2 ,$$

and

$$E(MSB) = \sigma^2 + \frac{(1-\delta_\beta)}{p(q-1)} \sum_{i,j}^{p,q} n_{ij} \beta_{j(i)}^2 + \frac{\delta_\beta}{p(q-1)} \left[N - \frac{1}{N} \sum_{j=1}^{q} n_{\cdot j}^2\right] \sigma_\beta^2 .$$

Finally, by the theory of Chapter II, we know that SSE is distributed as $\chi^2\sigma^2$ with N-pq d.f. . Hence

$$E\left(\frac{SSE}{\sigma^2}\right) = N\text{-}pq \quad \text{and} \quad E(MSE) = \sigma^2 .$$

These results are summarized in table 5.1, where

$$a = N - \frac{1}{N} \sum_{i=1}^{p} n_{i\cdot}^2 , \quad b = N - \frac{1}{N} \sum_{j=1}^{q} n_{\cdot j}^2 ,$$

δ_γ , δ_β are zero if the γ_i's , $\beta_{j(i)}$'s come from Type I populations and are one otherwise.

The following formulas are more convenient for computation:

Table 5.1

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Classes	p-1	$SST = \sum_{i=1}^{p} n_{i.}(\bar{Y}_{i..} - \bar{Y}_{...})^2$	$MST = \frac{SST}{p-1}$	$\sigma^2 + \frac{d_\beta\, a}{(p-1)N^2}\left[\sum_{j=1}^{q} n_{.j}^2 - \frac{N^2}{Q}\right]\sigma_\beta^2 + \frac{d_\tau\, a}{(p-1)}\sigma_\tau^2 + \frac{(1-d_\tau)}{p-1}\sum_{i=1}^{p} n_{i.}\tau_i^2$.
Subclasses	p(q-1)	$SSB = \sum_{i,j}^{p,q} n_{ij}(\bar{Y}_{ij.} - \bar{Y}_{i..})^2$	$MSB = \frac{SSB}{p(q-1)}$	$\sigma^2 + d_\beta \frac{b}{p(q-1)}\sigma_\beta^2 + \frac{(1-d_\beta)}{p(q-1)}\sum_{i,j}^{p,q} n_{ij}\beta_{j(i)}^2$
Error	N-pq	$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})^2$	$MSE = \frac{SSE}{N-pq}$	σ^2
Total	N-1	$\sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{...})^2$		

$$SST = \sum_{i=1}^{p} n_{i.}(\bar{Y}_{i..} - \bar{Y}_{...})^2 = \sum_{i=1}^{p} n_{i.}\bar{Y}_{i..}^2 - N\bar{Y}_{...}^2 = \sum_{i=1}^{p} \frac{Y_{i..}^2}{n_{i.}} - \frac{Y_{...}^2}{N} ,$$

$$SSB = \sum_{i,j}^{p,q} n_{ij}(\bar{Y}_{ij.} - \bar{Y}_{i..})^2 = \sum_{i,j}^{p,q} n_{ij}\bar{Y}_{ij.}^2 - \sum_{i=1}^{p} n_{i.}\bar{Y}_{i..}^2 = \sum_{i,j}^{p,q} \frac{Y_{ij.}^2}{n_{ij}} - \sum_{i=1}^{p} \frac{Y_{i..}^2}{n_{i.}} ,$$

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})^2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} Y_{ijk_{ij}}^2 - \sum_{i,j}^{p,q} \frac{Y_{ij.}^2}{n_{ij}} ,$$

and

$$\sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{...})^2 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} Y_{ijk_{ij}} - \frac{Y_{...}^2}{N} .$$

If $n_{ij} = 1$ $(i = 1,2,\ldots,p; j = 1,2,\ldots,q)$, $\bar{Y}_{ij.} = Y_{ijk_{ij}}$, SSE $= 0$, and it is impossible to carry out any of the F tests involving SSE.

If $n_{ij} = n$, $n_{i.} = qn$, $n_{.j} = pn$, $N = pqn$,

$$a = npq - \frac{pq^2n^2}{pqn} = qn(p-1), \quad b = pn(q-1),$$

$$\frac{1}{N^2}\left[\sum_{j=1}^{q} n_{.j}^2 - \frac{N^2}{Q}\right] = \frac{1}{N^2}\left[p^2qn^2 - \frac{p^2q^2n^2}{Q}\right] = \frac{pn}{N}\left(1 - \frac{q}{Q}\right) ,$$

$$\frac{1}{p-1}\sum_{i=1}^{p} n_{i.}\tau_i^2 = \frac{qn}{p-1}\sum_{i=1}^{p}\tau_i^2 \text{ and } \frac{1}{p(q-1)}\sum_{i,j}^{p,q} n_{ij}\beta_{j(i)}^2 = \frac{n}{p(q-1)}\sum_{i,j}^{p,q}\beta_{j(i)}^2 .$$

If we define

$$\sigma_\tau^2 = \frac{1}{p-1}\sum_{i=1}^{p}\tau_i^2 , \quad \sigma_\beta^2 = \frac{1}{p(q-1)}\sum_{i,j}^{p,q}\beta_{j(i)}^2 ,$$

when $n_{ij} = n$ for the Type I populations, we obtain table 5.2,

TABLE 5.2

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Classes	p-1	$SST = qn\sum_{i=1}^{p}(\bar{Y}_{i..}-\bar{Y}_{...})^2$	$MST = \frac{SST}{p-1}$	$\sigma^2 + n\left(1-\frac{q}{Q}\right)\sigma_\beta^2 + qn\,\sigma_\tau^2$
Subclasses	p(q-1)	$SSB = n\sum_{i,j}^{p,q}(\bar{Y}_{ij.}-\bar{Y}_{i..})^2$	$MSB = \frac{SSB}{p(q-1)}$	$\sigma^2 + n\,\sigma_\beta^2$
Error	N-pq	$SSE = \sum_{i,j,k}^{p,q,n}(Y_{ijk}-\bar{Y}_{ij.})^2$	$MSE = \frac{SSE}{N-pq}$	σ^2
Total	N-1	$\sum_{i,j,k}^{p,q,n}(Y_{ijk}-\bar{Y}_{...})^2$		

where we omit δ_β since $1-\frac{q}{Q}=0$ when $Q=q$.

5.5 Distributions of the Sums of Squares

Corresponding to the hypotheses

H_1: $\beta_{j(i)} = 0 \quad (i=1,2,\ldots,p; j=1,2,\ldots q-1)$

H_2: $\tau_i = 0 \quad (i=1,2,\ldots,p-1)$

for the Type I model, we have the hypotheses

$$\sigma_\beta = 0 \quad , \quad \sigma_\tau = 0 \quad ,$$

if the corresponding variables are from other than a Type I population. Since the populations have zero means, the corresponding variables are then equal to zero.

In the case of the Type I model, the theory of Chapter II implies that

$$t_i = \bar{Y}_{i..} - \bar{Y}_{...} \quad \text{and} , \; b_{j(i)} = \bar{Y}_{ij.} - \bar{Y}_{i..} \quad ,$$

$(i=1,2,\ldots,p; j=1,2,\ldots,q-1)$ are distributed independently of SSE. Therefore any function of these statistics is distributed independently of SSE and, in particular, this holds for t_p and $b_{q(i)}$ (i 1,2,...,p). These results were obtained for the model

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk_{ij}}$$

and they hold for the particular case where $Y_{ijk_{ij}} = \varepsilon_{ijk_{ij}}$.

Hence

$$\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...} \quad \text{and} \quad \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..} \qquad (i=1,2,\ldots,p; j=1,2,\ldots,q)$$

are distributed independently of SSE.

We shall now show that these expressions in the $\varepsilon_{ijk_{ij}}$'s are independent of each other. Since they have normal distributions, it is sufficient to prove that

$$\mathrm{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}\,,\ \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}) = 0 \quad .$$

Since

$$E(\bar{\varepsilon}_{i..}\,\bar{\varepsilon}_{ij.}) = \frac{n_{ij}}{n_{i.}n_{ij}}\,\sigma^2 = \frac{\sigma^2}{n_{i.}} \quad ,$$

$$E(\bar{\varepsilon}_{i..}^{\,2}) = \frac{\sigma^2}{n_{i.}} \quad ,$$

$$E(\bar{\varepsilon}_{.j.}\,\bar{\varepsilon}_{...}) = \frac{1}{n_{.j}N}\,n_{.j}\,\sigma^2 = \frac{\sigma^2}{N} \quad ,$$

$$E(\bar{\varepsilon}_{i..}\,\bar{\varepsilon}_{...}) = \frac{1}{n_{i.}N}\,n_{i.}\,\sigma^2 = \frac{\sigma^2}{N} \quad ,$$

we have

$$\mathrm{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}\,,\ \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}) = \sigma^2\left(\frac{1}{n_{i.}} - \frac{1}{n_{i.}} - \frac{1}{N} + \frac{1}{N}\right) = 0 \quad .$$

In section 5.1 we tested the hypotheses H_1 and H_2 for the Type I model by the statistics

$$F_1 = \frac{MSB}{MSE} \quad \text{and} \quad F_2 = \frac{MST}{MSE} \quad ,$$

respectively, since subject to the corresponding hypotheses,

$$SST = \sum_{i=1}^{p} n_{i.}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2 , \quad SSB = \sum_{i,j}^{p,q} n_{ij}(\bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..})^2 ,$$

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (\varepsilon_{ijk_{ij}} - \bar{\varepsilon}_{ij.})^2 \quad ,$$

are independently distributed as $\chi^2\sigma^2$ with p-1, p(q-1) and N-pq d.f. , respectively.

We shall now determine what tests can be made when we are not dealing with a Type I model. In section 5.4 we saw that

$$SST = \sum_{i=1}^{p} n_{i\cdot}[\tau_i - \bar{\tau}_\cdot + \bar{\beta}_{\cdot(i)} - \bar{\beta}_{\cdot(\cdot)} + \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot\cdot\cdot}]^2$$

and

$$SSB = \sum_{ij}^{p,q} n_{ij}[\beta_{j(i)} - \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot}]^2 \quad .$$

Subject to H_1, the sum of squares, SSB, reduces to the corresponding expression for the Type I model and the same test applies.

Our problem is now reduced to testing H_2 when we are not dealing with a Type I model. Suppose we first consider the Type II model where the τ_i's and $\beta_{j(i)}$'s are NID with zero means and variances σ_τ^2 and σ_β^2 respectively. When H_2: $\sigma_\tau = 0$ holds, let

$$y_i = \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{i\cdot\cdot} = \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}\beta_{j(i)} + \bar{\varepsilon}_{i\cdot\cdot} \quad .$$

Then

$$\bar{y}_\cdot = \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}y_i = \frac{1}{N^2}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{i\cdot}n_{\cdot j}\beta_{j(i)} + \frac{1}{N}\sum_{i=1}^{p} n_{i\cdot}\bar{\varepsilon}_{i\cdot\cdot}$$

$$= \frac{1}{N}\sum_{ij}^{p,q} n_{ij}\beta_{j(i)} + \bar{\varepsilon}_{\cdot\cdot\cdot} = \bar{\beta}_{\cdot(\cdot)} + \bar{\varepsilon}_{\cdot\cdot\cdot} \quad ,$$

and

$$SST = \sum_{i=1}^{p} n_{i\cdot}(y_i - \bar{y}_\cdot)^2$$

where

$$E(y_i) = 0 \quad , \quad Var(y_i) = \frac{1}{N^2}\sum_{j=1}^{q} n_{\cdot j}^2\,\sigma_\beta^2 \; ,$$

$$cov(y_i, y_{i'}) = 0 .$$

With similar situations in Chapter III and Chapter IV we concluded that

$$SST/E(MST)$$

did not, in general, have a χ^2 distribution unless $n_{ij} = n$. Assuming, therefore, that $n_{ij} = n$, we find that

$$n_{i.} = qn, \quad n_{.j} = pn, \quad N = pqn,$$

$$\bar{\beta}_{\cdot(i)} = \frac{1}{N}\sum_{j=1}^{q} n_{\cdot j}\,\beta_{j(i)} = \frac{1}{q}\sum_{j=1}^{q}\beta_{j(i)} , \quad \bar{\beta}_{\cdot(\cdot)} = \frac{1}{pq}\sum_{ij}^{p,q}\beta_{j(i)} .$$

Without imposing the restriction that H_2 holds, let

$$y_i = \tau_i + \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{i\cdot\cdot} ,$$

$$\bar{y}_. = \frac{1}{p}\sum_{i=1}^{p}(\tau_i + \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{i\cdot\cdot}) = \bar{\tau}_. + \bar{\beta}_{\cdot(\cdot)} + \bar{\varepsilon}_{\cdots} .$$

Then

$$E(y_i) = 0 , \quad Var(y_i) = \sigma_\tau^2 + \frac{\sigma_\beta^2}{q} + \frac{\sigma^2}{qn} ,$$

$$cov(y_i, y_{i'}) = 0 ,$$

$$SST = \sum_{i=1}^{p} qn(y_i - \bar{y}_.)^2 ,$$

and by section 5.4,

$$E(MST) = \sigma^2 + n\sigma_\beta^2 + qn\sigma_\tau^2 .$$

Hence

$$\frac{SST}{E(MST} = \frac{\sum_{i=1}^{p}(y_i - \bar{y}_.)^2}{\sigma_\tau^2 + \frac{\sigma_\beta^2}{q} + \frac{\sigma^2}{qn}} ,$$

by Chapter II has a χ^2 distribution with p-1 d.f. .

Consider the two sets of variables

$$\bar{\beta}_{\cdot(i)} - \bar{\beta}_{\cdot(\cdot)} \ , \quad \beta_{j(i)} - \bar{\beta}_{\cdot(i)} \ .$$

We have

$$\begin{aligned} \operatorname{cov}(\bar{\beta}_{\cdot(i)} - \bar{\beta}_{\cdot(\cdot)}, \beta_{j(i)} - \bar{\beta}_{\cdot(i)}) &= E(\bar{\beta}_{\cdot(i)}\beta_{j(i)}) - E(\bar{\beta}^2_{\cdot(i)}) \\ &\quad - E(\bar{\beta}_{\cdot(\cdot)}\beta_{j(i)}) + E(\bar{\beta}_{\cdot(\cdot)}\bar{\beta}_{\cdot(i)}) \\ &= \frac{1}{q}\sum_{j'=1}^{q} E(\beta_{j'(i)}\beta_{j(i)}) - \frac{1}{q^2}\sum_{j,j'}^{q} E(\beta_{j(i)}\beta_{j'(i)}) \\ &\quad - \frac{1}{pq}\sum_{i',j'}^{p,q} E(\beta_{j'(i')}\beta_{j(i)}) + \frac{1}{pq^2}\sum_{i',j,j'}^{p,q} E(\beta_{j'(i')}\beta_{j(i)}) \\ &= \frac{\sigma_\beta^2}{q} - \frac{\sigma_\beta^2}{q} - \frac{\sigma_\beta^2}{pq} + \frac{\sigma_\beta^2}{pq} = 0 \ . \end{aligned}$$

Hence, the two sets are independent. Since we have already proved the corresponding relations for the $\bar{\varepsilon}_{ij\cdot}$'s , it follows that the two sets of variables

$$\tau_i - \bar{\tau}_{\cdot} + \bar{\beta}_{\cdot(i)} - \bar{\beta}_{\cdot(\cdot)} + \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot\cdot\cdot} \ ,$$

$$\beta_{j(i)} - \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} \ ,$$

are independently distributed and hence so are SST and SSB.

For the Type II model we display the above results in table 5.3:

TABLE 5.3

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Classes	p-1	SST	MST	$\sigma^2 + n\sigma_\beta^2 + qn\sigma_\tau^2$
Subclasses	p(q-1)	SSB	MSB	$\sigma^2 + n\sigma_\beta^2$
Error	N-pq	SSE	MSE	σ^2
Total	N-1			

From the last column of this table we can determine the tests for H_1 and H_2. We know that the sums of squares divided by the expected value of their mean squares, are independently distributed as χ^2. If we also divide by the corresponding degrees of freedom, the ratio of any two has the F distribution. As we have seen in the previous chapter, the computation can be carried out only when the expected mean squares cancel out in this ratio. Thus we have

$$F_1 = \frac{MSB}{MSE} \quad , \qquad F_2 = \frac{MST}{MSB}$$

to test the hypotheses H_1 and H_2 respectively. Both of these tests can be carried out when $n_{ij} = n$ and when this condition does not hold only the test for H_1 can be carried out. The test for H_2 differs from the corresponding test for the Type I model, the denominator now being MSB rather than MSE.

Finally we shall consider the Type III model. First, we have already seen that, subject to H_1, the same test applies here as for the Type I model. We can not expect to obtain χ^2 distributions necessary for an F test of H_2 since such a test depends on the $\beta_{j(i)}$'s being normally distributed unless the terms involving the $\beta_{j(i)}$'s reduce to zero as in the Type I model. Approximate tests have been established to enable one to carry out the F test subject to H_2.[1,2]

[1]Satterthwaite, F.E., An Approximate Distribution of Estimates of Variance Components. Biometrics Bull, 2 (1946), pp.110-114

[2]Welch, B.L., The Specification of Rules for Rejecting too Variable a Product. J. Royal Stat. Soc. (Supp.), 3 (1936), p. 29

CHAPTER VI

LATIN SQUARE MODELS

6.1 The Type I Model for an $m \times m$ Latin Square

We shall now consider an experiment that has been designed in a manner such that each treatment is assigned at random within a row and a column so that all treatments appear once in each row and column. The regression model for this design is

$$Y_{ijk} = \mu + \alpha_i + \gamma_j + \tau_k + \varepsilon_{ijk} \quad ,$$
$$(i,j,k = 1,2,\ldots,m)$$

where α_i , γ_j , and τ_k are called the row, column, and treatment effects respectively. Once i and j are specified we know k. Hence k is a function of i and j. The parameters are subject to the restrictions

$$\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{m} \gamma_j = \sum_{k=1}^{m} \tau_k = 0$$

The analysis of this design has been carried out by Bedrosian[1] and the results are exhibited in table 6.1, where $\bar{Y}_{i..}$, $\bar{Y}_{.j.}$, and $\bar{Y}_{..k}$ are the means of the ith row, jth column, and all observations on the kth treatment respectively.

We shall be interested in testing the three hypotheses

H_1: $\alpha_i = 0 \quad (i=1,2,\ldots,m)$,

H_2: $\gamma_j = 0 \quad (j=1,2,\ldots,m)$,

---

[1] Bedrosian, Peter, Orthogonal Latin Squares and Incomplete Balanced Block Designs. Hamilton: unpublished thesis, 1953, p.90.

TABLE 6.1

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Rows	m-1	$SSA = m\sum_{i=1}^{m}(\bar{Y}_{i..}-\bar{Y}_{...})^2$	MSA = SSA/m-1
Columns	m-1	$SSC = m\sum_{j=1}^{m}(\bar{Y}_{.j.}-\bar{Y}_{...})^2$	MSC = SSC/m-1
Treatments	m-1	$SST = m\sum_{k=1}^{m}(\bar{Y}_{..k}-\bar{Y}_{...})^2$	MST = SST/m-1
Error	(m-1)(m-2)	$SSE = \sum_{i,j}^{m}(Y_{ijk}-\bar{Y}_{i..}-\bar{Y}_{.j.}-\bar{Y}_{..k}+2\bar{Y}_{...})^2$	MSE = SSE/(m-1)(m-2)
Total	m^2-1	$\sum_{i,j}^{m}(Y_{ijk}-\bar{Y}_{...})^2$	

H_3: $\tau_k = 0 \quad (k=1,2,\ldots,m)$.

The theory of Chapter II tells us that, subject to the corresponding hypotheses, SSA, SSC, SST, and SSE are independently distributed as $\chi^2\sigma^2$, the first three sums with m-1 d.f. each and the last sum with (m-1)(m-2) d.f. . The hypotheses H_1, H_2, and H_3 are tested by the statistics

$$F_1 = \frac{MSA}{MSE} \quad , \quad F_2 = \frac{MSC}{MSE} \quad , \quad F_3 = \frac{MST}{MSE} \quad ,$$

respectively.

6.2 Other Models for the m×m Latin Square

We have that

$$Y_{ijk} = \mu + \alpha_i + \gamma_j + \tau_k + \varepsilon_{ijk}$$

For the Type II model we assume that the α_i's , γ_j's, τ_k's and ε_{ijk}'s are NID with zero means and variances σ_α^2, σ_γ^2, σ_τ^2, and σ^2 respectively. Then we have

$$E(Y_{ijk}) = \mu$$

and

$$Var(Y_{ijk}) = \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\tau^2 + \sigma^2$$

For the Type III model we assume that the α_i's, γ_j's , and τ_k's come from finite independent populations of size M_α, M_γ, and M_τ, say, respectively, with zero means and variances

$$\sigma_\alpha^2 = \frac{1}{M_\alpha - 1}\sum_{i=1}^{M_\alpha}\alpha_i^2 \quad , \quad \sigma_\gamma^2 = \frac{1}{M_\gamma - 1}\sum_{j=1}^{M_\gamma}\gamma_j^2 \quad ,$$

$$\sigma_\tau^2 = \frac{1}{M_\tau - 1}\sum_{k=1}^{M_\tau}\tau_k^2 \quad .$$

The assumption of zero means implies that

$$\sum_{i=1}^{M_\alpha} \alpha_i = 0 \ , \ \sum_{j=1}^{M_\gamma} \gamma_j = 0 \ , \ \sum_{k=1}^{M_\tau} \tau_k = 0 \ .$$

For the mixed model we may assume that the α_i's , γ_j's , and τ_k's are of any of the types described above.

6.3 The Expected Values of the Sums of Squares for an $m \times m$ Latin Square

In every case we shall arbitrarily begin with the sums of squares obtained for the Type I model. Note that for i fixed, as j goes from 1 to m, k goes from 1 to m, although not necessarily in that order. Similarily for j fixed, k goes from 1 to m as i goes from 1 to m. Finally by $Y_{..k}$ we shall mean the sum over all values of i, j, corresponding to the κth treatment, while $Y_{...}$ is the sum over all values of i, j, giving all values of k. Then, since

$$Y_{ijk} = \mu + \alpha_i + \gamma_j + \tau_k + \varepsilon_{ijk} \ ,$$

$$\bar{Y}_{i..} = \mu + \alpha_i + \bar{\gamma}_. + \bar{\tau}_. + \bar{\varepsilon}_{i..} \ ,$$

$$\bar{Y}_{.j.} = \mu + \bar{\alpha}_. + \gamma_j + \bar{\tau}_. + \bar{\varepsilon}_{.j.} \ ,$$

$$\bar{Y}_{..k} = \mu + \bar{\alpha}_. + \bar{\gamma}_. + \tau_k + \bar{\varepsilon}_{..k} \ ,$$

$$\bar{Y}_{...} = \mu + \bar{\alpha}_. + \bar{\gamma}_. + \bar{\tau}_. + \bar{\varepsilon}_{...} \ ,$$

where

$$\bar{\alpha}_. = \frac{1}{m}\sum_{i=1}^{m} \alpha_i \ , \quad \bar{\gamma}_. = \frac{1}{m}\sum_{j=1}^{m} \gamma_j \ , \quad \bar{\tau}_. = \frac{1}{m}\sum_{k=1}^{m} \tau_k \ ,$$

$$SSA = m\sum_{i=1}^{m}(\bar{Y}_{i..} - \bar{Y}_{...})^2 = m\sum_{i=1}^{m}(\alpha_i - \bar{\alpha}_. + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2,$$

$$SSC = m\sum_{j=1}^{m}(\bar{Y}_{.j.} - \bar{Y}_{...})^2 = m\sum_{j=1}^{m}(\gamma_j - \bar{\gamma}_. + \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...})^2,$$

$$SST = m\sum_{k=1}^{m}(\bar{Y}_{..k} - \bar{Y}_{...})^2 = m\sum_{k=1}^{m}(\tau_k - \bar{\tau}_. + \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})^2,$$

$$SSE = \sum_{i,j}^{m}(Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}_{...})^2$$

$$= \sum_{i,j}^{m}(\varepsilon_{ijk} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{..k} + 2\bar{\varepsilon}_{...})^2 .$$

To evaluate the expected values of these sums of squares by Theorem 3.2, we shall need the variances and covariances of the following sets of variables,

$$\alpha_i + \bar{\varepsilon}_{i..} \quad , \quad \gamma_j + \bar{\varepsilon}_{.j.} \quad , \quad \tau_k + \bar{\varepsilon}_{..k} \quad .$$

Therefore we must compute

$$E(\alpha_i^2),\ E(\alpha_i\,\alpha_{i'}),\ E(\gamma_j^2),\ E(\gamma_j\,\gamma_{j'}),\ E(\tau_k^2) \text{ and } E(\tau_k\,\tau_{k'})$$

in a form which will be valid regardless of the nature of the populations used.

If α_i comes from a Type II population, $E(\alpha_i^2) = \sigma_\alpha^2$, and if it comes from a Type III population,

$$E(\alpha_i^2) = \frac{1}{M_\alpha}\sum_{i=1}^{M_\alpha}\alpha_i^2 = \left(1 - \frac{1}{M_\alpha}\right)\sigma_\alpha^2 ,$$

which gives the formula for the Type II case if we let $M_\alpha \to \infty$. For the Type II case, $E(\alpha_i, \alpha_{i'}) = 0$, and for the Type III case

$$E(\alpha_i\,\alpha_{i'}) = \sum_{i \neq i'}^{M_\alpha}\frac{\alpha_i\,\alpha_{i'}}{M_\alpha(M_\alpha - 1)} = -\sum_{i=1}^{M_\alpha}\frac{\alpha_i^2}{M_\alpha(M_\alpha - 1)} = -\frac{\sigma_\alpha^2}{M_\alpha}$$

since

$$0 = \sum_{i,i'}^{M_\alpha} \alpha_i \alpha_{i'} = \sum_{i=1}^{M_\alpha} \alpha_i^2 + \sum_{i \neq i'}^{M_\alpha} \alpha_i \alpha_{i'} \quad .$$

Thus the formula for the Type II case is again included in the Type III case. Similarly for the γ_j's and τ_k's ,

$$E(\gamma_j^2) = \left(1 - \frac{1}{M_\gamma}\right)\sigma_\gamma^2 \,, \quad E(\gamma_j \gamma_{j'}) = -\frac{\sigma_\gamma^2}{M_\gamma} \,,$$

and

$$E(\tau_k^2) = \left(1 - \frac{1}{M_\tau}\right)\sigma_\tau^2 \,, \quad E(\tau_k \tau_{k'}) = -\frac{\sigma_\tau^2}{M_\tau}$$

To evaluate E(SSA) we let

$$y_i = \alpha_i + \bar{\varepsilon}_{i..} \quad , \quad \bar{y}_. = \frac{1}{m}\sum_{i=1}^{m} y_i = \bar{\alpha}_. + \bar{\varepsilon}_{...} \quad ,$$

so that

$$SSA = m \sum_{i=1}^{m} (y_i - \bar{y}_.)^2 \,,$$

$$\mu_i = E(y_i) = E(\alpha_i) = (1 - \delta_\alpha)\alpha_i \quad ,$$

$$\bar{\mu}_. = \frac{1}{m}\sum_{i=1}^{m} \mu_i = 0$$

where $\delta_\alpha = 0$ if the α_i's come from a Type I population and $\delta_\alpha = 1$ otherwise. Also

$$\sigma_i^2 = Var(y_i) = E[\alpha_i + \bar{\varepsilon}_{i..} - E(\alpha_i)]^2$$

$$= E[\alpha_i - E(\alpha_i)]^2 + E(\bar{\varepsilon}_{i..}^2)$$

$$= \delta_\alpha \left(1 - \frac{1}{M_\alpha}\right)\sigma_\alpha^2 + \frac{\sigma^2}{m} \quad ,$$

and in order to use corollary 3.21 we find that

$$\bar{\sigma}_{.}^{2} = \frac{1}{m}\sum_{i=1}^{m} \sigma_i^2 = \delta_\alpha\left(1-\frac{1}{M_\alpha}\right)\sigma_\alpha^2 + \frac{\sigma^2}{m} \quad .$$

Next

$$\lambda = \operatorname{cov}(y_i, y_{i'}) = E\left\{\left[\alpha_i - E(\alpha_i)\right]\left[\alpha_{i'} - E(\alpha_{i'})\right]\right\}$$

$$= -\delta_\alpha \frac{\sigma_\alpha^2}{M_\alpha} \quad ,$$

$$\bar{\sigma}_{.}^{2} - \lambda = \delta_\alpha \sigma_\alpha^2 + \frac{\sigma^2}{m} \quad ,$$

$$E(SSA) = (1-\delta_\alpha)\, m \sum_{i=1}^{m} \alpha_i^2 + m(m-1)\left(\delta_\alpha \sigma_\alpha^2 + \frac{\sigma^2}{m}\right)$$

$$=(m-1)\sigma^2 + \delta_\alpha m(m-1)\sigma_\alpha^2 + (1-\delta_\alpha)\, m \sum_{i=1}^{m} \alpha_i^2$$

and

$$E(MSA) = \sigma^2 + \delta_\alpha m \sigma_\alpha^2 + (1-\delta_\alpha)\frac{m}{m-1}\sum_{i=1}^{m}\alpha_i^2 \quad .$$

Similarly

$$E(MSC) = \sigma^2 + \delta_\gamma m \sigma_\gamma^2 + (1-\delta_\gamma)\frac{m}{m-1}\sum_{j=1}^{m}\gamma_j^2 \quad ,$$

$$E(MST) = \sigma^2 + \delta_\tau m \sigma_\tau^2 + (1-\delta_\tau)\frac{m}{m-1}\sum_{k=1}^{m}\tau_k^2 \quad .$$

Finally, by the theory of Chapter II, we know that SSE is distributed as $\chi^2\sigma^2$ with (m-1)(m-2) d.f. . Hence

$$E\left(\frac{SSE}{\sigma^2}\right) = (m-1)(m-2)$$

and

$$E(MSE) = \sigma^2 \quad .$$

If we define

$$\sigma_\alpha^2 = \frac{1}{m-1}\sum_{i=1}^{m}\alpha_i^2 \;, \quad \sigma_\gamma^2 = \frac{1}{m-1}\sum_{j=1}^{m}\gamma_j^2 \;, \quad \sigma_\tau^2 = \frac{1}{m-1}\sum_{k=1}^{m}\tau_k^2 \;,$$

for the Type I populations, the above results are summarized in table 6.2.

The following formulas are more convenient for computation:

$$SSA = m\sum_{i=1}^{m}(\bar{Y}_{i..} - \bar{Y}_{...})^2 = \sum_{i=1}^{m}\frac{Y_{i..}^2}{m} - \frac{Y_{...}^2}{m^2} \;,$$

$$SSC = m\sum_{j=1}^{m}(\bar{Y}_{.j.} - \bar{Y}_{...})^2 = \sum_{j=1}^{m}\frac{Y_{.j.}^2}{m} - \frac{Y_{...}^2}{m^2} \;,$$

$$SST = m\sum_{k=1}^{m}(\bar{Y}_{..k} - \bar{Y}_{...})^2 = \sum_{k=1}^{m}\frac{Y_{..k}^2}{m} - \frac{Y_{...}^2}{m^2} \;,$$

$$\sum_{i,j}^{m}(Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i,j}^{m} Y_{ijk}^2 - \frac{Y_{...}^2}{m^2} \;,$$

and SSE is obtained by subtraction.

6.4 Distributions of the Sums of Squares for an $m \times m$ Latin Square

Corresponding to the hypotheses

H_1: $\alpha_i = 0$, H_2: $\gamma_j = 0$, H_3: $\tau_k = 0$,

we have the hypotheses

$$\sigma_\alpha = 0 \;, \quad \sigma_\gamma = 0 \;, \quad \sigma_\tau = 0 \;,$$

if the corresponding variables are from other than a Type I population. Then, since the populations have zero means it follows that the corresponding variables are equal to zero.

TABLE 6.2

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Rows	m-1	$SSA = m \sum_{i=1}^{m} (\bar{Y}_{i..} - \bar{Y}_{...})^2$	$MSA = \frac{SSA}{m-1}$	$\sigma^2 + m\sigma_\alpha^2$
Columns	m-1	$SSC = m \sum_{j=1}^{m} (\bar{Y}_{.j.} - \bar{Y}_{...})^2$	$MSC = \frac{SSC}{m-1}$	$\sigma^2 + m\sigma_\gamma^2$
Treatments	m-1	$SST = m \sum_{k=1}^{m} (\bar{Y}_{..k} - \bar{Y}_{...})^2$	$MST = \frac{SST}{m-1}$	$\sigma^2 + m\sigma_\tau^2$
Error	(m-1)(m-2)	$SSE = \sum_{i,j} (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}_{...})^2$	$MSE = \frac{SSE}{(m-1)(m-2)}$	σ^2
Total	m^2-1	$\sum_{i,j}^{m} (Y_{ijk} - \bar{Y}_{...})^2$		

In the case of the Type I model, the theory of Chapter II implies that

$\bar{Y}_{i..} - \bar{Y}_{...}$, $\bar{Y}_{.j.} - \bar{Y}_{...}$, $\bar{Y}_{..k} - \bar{Y}_{...}$, $(i,j,k=1,2,\ldots,m-1)$

are distributed independently of SSE. Therefore any function of these expressions is distributed independently of SSE and, in particular, this holds for $\bar{Y}_{m..} - \bar{Y}_{...}$, $\bar{Y}_{.m.} - \bar{Y}_{...}$, and $\bar{Y}_{..m} - \bar{Y}_{...}$. These results were obtained for the model

$$Y_{ijk} = \mu + \alpha_i + \gamma_j + \tau_k + \varepsilon_{ijk}$$

and they hold for the particular case where $Y_{ijk} = \varepsilon_{ijk}$. Hence

$$\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...} \;,\quad \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...} \quad \text{and} \quad \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...}$$

are distributed independently of SSE.

We shall now show that these expressions in the ε_{ijk}'s are independent of each other. Since they have normal distributions, it is sufficient to prove that

$$\operatorname{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...} , \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...}) = \operatorname{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...} , \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})$$

$$= \operatorname{cov}(\bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...} , \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...}) = 0 \;.$$

Since

$$E(\bar{\varepsilon}_{i..}\,\bar{\varepsilon}_{.j.}) = \sigma^2/m^2 \;,$$

$$E(\bar{\varepsilon}_{i..}\,\bar{\varepsilon}_{...}) = \frac{m}{m^3}\sigma^2 = \sigma^2/m^2 \;,$$

$$E(\bar{\varepsilon}_{.j.}\,\bar{\varepsilon}_{...}) = \frac{m}{m^3}\sigma^2 = \sigma^2/m^2 \;,$$

$$E(\bar{\varepsilon}_{...}^2) = \sigma^2/m^2 \;,$$

we have

$$\operatorname{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}, \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...}) = \sigma^2\left(\frac{1}{m^2} - \frac{1}{m^2} - \frac{1}{m^2} + \frac{1}{m^2}\right) = 0 \ .$$

Similarly

$$\operatorname{cov}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}, \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...}) = \operatorname{cov}(\bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...}, \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...}) = 0 \ .$$

In section 6.1 we tested the hypotheses H_1, H_2, and H_3 for the Type I model by the statistics

$$F_1 = \frac{MSA}{MSE}, \qquad F_2 = \frac{MSC}{MSE}, \qquad F_3 = \frac{MST}{MSE},$$

respectively, since subject to the corresponding hypotheses,

$$SSA = m\sum_{i=1}^{m}(\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2 \ , \qquad SSC = m\sum_{j=1}^{m}(\bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...})^2 \ ,$$

$$SST = m\sum_{k=1}^{m}(\bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})^2 \ , \qquad SSE = \sum_{i,j}^{m}(\varepsilon_{ijk} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{..k} + 2\bar{\varepsilon}_{...})^2$$

are independently distributed as $\chi^2\sigma^2$, the first three sums with m-1 d.f. each and the last sum with (m-1)(m-2) d.f. .

We shall now determine what tests can be made when we are not dealing with a Type I model. In section 6.3 we saw that

$$SSA = m\sum_{i=1}^{m}(\alpha_i - \bar{\alpha}_{.} + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2$$

$$SSC = m\sum_{j=1}^{m}(\gamma_j - \bar{\gamma}_{.} + \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...})^2$$

and

$$SST = m\sum_{k=1}^{m}(\tau_k - \bar{\tau}_{.} + \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})^2 \ .$$

Subject to H_1, H_2, and H_3, the sums of squares SSA, SSC, and SST reduce to the corresponding expressions for the Type I model and the same tests apply.

6.5 The Type I Model for Replicated Latin Squares

We shall now carry through the above theory when we have r identical Latin squares. Denoting the observations of the ℓth Latin square by $Y_{ijk}^{(\ell)}$, the regression model is

$$Y_{ijk}^{(\ell)} = \mu + \alpha_i^{(\ell)} + \gamma_j^{(\ell)} + \tau_k + \rho_{(\ell)} + \varepsilon_{ijk}^{(\ell)} \quad ,$$
$$(i,j,k = 1,2,\ldots,m;\ \ell = 1,2,\ldots,r)$$

where $\alpha_i^{(\ell)}$, $\gamma_j^{(\ell)}$, τ_k, $\rho_{(\ell)}$, are called the row, column, treatment, and replicate effects of the ℓth Latin square, respectively. The parameters are subject to the restrictions

$$\sum_{i=1}^{m} \alpha_i^{(\ell)} = \sum_{j=1}^{m} \gamma_i^{(\ell)} = \sum_{k=1}^{m} \tau_k = \sum_{\ell=1}^{r} \rho_{(\ell)} = 0 \quad (\ell = 1,2,\cdots,r) .$$

Again we exhibit the results from Bedrosian's thesis in table 6.3, where $\bar{Y}_{i..}^{(\ell)}$, $\bar{Y}_{.j.}^{(\ell)}$ are the means of the ith row and jth column of the ℓth Latin square; $\bar{Y}_{..k}$ is the mean of all the observations of the kth treatment over all the Latin squares; $\bar{Y}_{...}^{(\ell)}$ is the mean of all observations in the ℓth Latin square and $\bar{Y}_{...}$ is the grand mean.

We shall be interested in testing the four hypotheses

H_1: $\alpha_i^{(\ell)} = 0 \quad (i = 1,2,\ldots,m;\ \ell = 1,2,\ldots,r)$,

H_2: $\gamma_j^{(\ell)} = 0 \quad (j = 1,2,\ldots,m;\ \ell = 1,2,\ldots,r)$,

H_3: $\tau_k = 0 \quad (k = 1,2,\ldots,m)$,

H_4: $\rho_{(\ell)} = 0 \quad (\ell = 1,2,\ldots,r)$.

TABLE 6.3

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Rows	r(m-1)	$SSA = m\sum_{\ell=1}^{r}\sum_{i=1}^{m}\left(\bar{Y}_{i..}^{(\ell)} - \bar{Y}_{...}^{(\ell)}\right)^2$	$MSA = \frac{SSA}{r(m-1)}$
Columns	r(m-1)	$SSC = m\sum_{\ell=1}^{r}\sum_{j=1}^{m}\left(\bar{Y}_{.j.}^{(\ell)} - \bar{Y}_{...}^{(\ell)}\right)^2$	$MSC = \frac{SSC}{r(m-1)}$
Treatments	m-1	$SST = mr\sum_{k=1}^{m}\left(\bar{Y}_{..k} - \bar{Y}_{...}\right)^2$	$MST = \frac{SST}{m-1}$
Replications	r-1	$SSP = m^2\sum_{\ell=1}^{r}\left(\bar{Y}_{...}^{(\ell)} - \bar{Y}_{...}\right)^2$	$MSP = \frac{SSP}{r-1}$
Error	(m-1)(mr-r-1)	$SSE = \sum_{\ell=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(Y_{ijk}^{(\ell)} - \bar{Y}_{i..}^{(\ell)} - \bar{Y}_{.j.}^{(\ell)} - \bar{Y}_{..k} + \bar{Y}_{...}^{(\ell)} + \bar{Y}_{...}\right)^2$	$MSE = \frac{SSE}{(m-1)(mr-r-1)}$
Total	m^2r-1	$\sum_{\ell=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(Y_{ijk}^{(\ell)} - \bar{Y}_{...}\right)^2$	

The theory of Chapter II tells us that, subject to the corresponding hypotheses, SSA, SSC, SST, SSP and SSE are independently distributed as $\chi^2\sigma^2$ with r(m-1), r(m-1), (m-1), (r-1), and (m-1)(mr-r-1) d.f. respectively. The hypotheses H_1, H_2, H_3, and H_4 are tested by the statistics

$$F_1 = \frac{MSA}{MSE} , \quad F_2 = \frac{MSC}{MSE} , \quad F_3 = \frac{MST}{MSE} , \quad F_4 = \frac{MSP}{MSE} ,$$

respectively.

6.6 Other Models for Replicated Latin Squares

Our regression model is

$$Y_{ijk}^{(\ell)} = \mu + \alpha_i^{(\ell)} + \gamma_j^{(\ell)} + \tau_k + \rho_{(\ell)} + \varepsilon_{ijk}^{(\ell)} .$$

For the Type II model we assume that the $\alpha_i^{(\ell)}$'s , $\gamma_j^{(\ell)}$'s , τ_k's , $\rho_{(\ell)}$'s , and $\varepsilon_{ijk}^{(\ell)}$'s are NID with zero means and variances $\sigma_{\alpha(\ell)}^2$, $\sigma_{\gamma(\ell)}^2$, σ_τ^2, σ_ρ^2 , and σ^2 respectively. Then we have

$$E(Y_{ijk}^{(\ell)}) = \mu$$

and

$$Var(Y_{ijk}^{(\ell)}) = \sigma_{\alpha(\ell)}^2 + \sigma_{\gamma(\ell)}^2 + \sigma_\tau^2 + \sigma_\rho^2 + \sigma^2$$

For the Type III model we assume that the $\alpha_i^{(\ell)}$'s , $\gamma_j^{(\ell)}$'s , τ_k's , and $\rho_{(\ell)}$'s come from finite independent populations of size $M_{\alpha(\ell)}$, $M_{\gamma(\ell)}$, M_τ and M_ρ , say, respectively, with zero means and variances

$$\sigma_{\alpha(\ell)}^2 = \frac{1}{M_{\alpha(\ell)}-1} \sum_{i=1}^{M_{\alpha(\ell)}} (\alpha_i^{(\ell)})^2, \quad \sigma_{\gamma(\ell)}^2 = \frac{1}{M_{\gamma(\ell)}-1} \sum_{j=1}^{M_{\gamma(\ell)}} (\gamma_j^{(\ell)})^2,$$

$$\sigma_\tau^2 = \frac{1}{M_\tau - 1} \sum_{k=1}^{M_\tau} \tau_k^2 \quad , \quad \sigma_\rho^2 = \frac{1}{M_\rho - 1} \sum_{\ell=1}^{M_\rho} \rho_{(\ell)}^2 \quad .$$

The assumption of zero means implies that

$$\sum_{i=1}^{M_{\alpha}(\ell)} \alpha_i^{(\ell)} = 0 \ , \ \sum_{j=1}^{M_{\gamma}(\ell)} \gamma_j^{(\ell)} = 0 \ , \ \sum_{k=1}^{M_\tau} \tau_k = 0 \ , \ \sum_{\ell=1}^{M_\rho} \rho_{(\ell)} = 0 \ .$$

For the mixed model we may assume that the $\alpha_i^{(\ell)}$'s , $\gamma_j^{(\ell)}$'s , τ_k's , and $\rho_{(\ell)}$'s are of any of the types described above.

6.7 The Expected Values of the Sums of Squares for Replicated Latin Squares

We shall arbitrarily begin with the sums of squares obtained for the Type I model and, since

$$Y_{ijk}^{(\ell)} = \mu + \alpha_i^{(\ell)} + \gamma_j^{(\ell)} + \tau_k + \rho_{(\ell)} + \varepsilon_{ijk}^{(\ell)} \quad ,$$

$$\bar{Y}_{i..}^{(\ell)} = \mu + \alpha_i^{(\ell)} + \bar{\gamma}_.^{(\ell)} + \bar{\tau}_. + \rho_{(\ell)} + \bar{\varepsilon}_{i..}^{(\ell)} \quad ,$$

$$\bar{Y}_{.j.}^{(\ell)} = \mu + \bar{\alpha}_.^{(\ell)} + \gamma_j^{(\ell)} + \bar{\tau}_. + \rho_{(\ell)} + \bar{\varepsilon}_{.j.}^{(\ell)} \quad ,$$

$$\bar{Y}_{..k} = \mu + \bar{\alpha}_. + \bar{\gamma}_. + \tau_k + \bar{\rho} + \bar{\varepsilon}_{..k} \quad ,$$

$$\bar{Y}_{...}^{(\ell)} = \mu + \bar{\alpha}_.^{(\ell)} + \bar{\gamma}_.^{(\ell)} + \bar{\tau}_. + \rho_{(\ell)} + \bar{\varepsilon}_{...}^{(\ell)} \quad ,$$

$$\bar{Y}_{...} = \mu + \bar{\alpha}_. + \bar{\gamma}_. + \bar{\tau}_. + \bar{\rho} + \bar{\varepsilon}_{...} \quad ,$$

where

$$\bar{\alpha}_.^{(\ell)} = \frac{1}{m} \sum_{i=1}^{m} \alpha_i^{(\ell)} , \quad \bar{\gamma}_.^{(\ell)} = \frac{1}{m} \sum_{j=1}^{m} \gamma_j^{(\ell)} \ , \quad \bar{\tau}_. = \frac{1}{m} \sum_{k=1}^{m} \tau_k \quad ,$$

$$\bar{\alpha}_{.} = \frac{1}{mr}\sum_{i,l}^{m,r} \alpha_i^{(l)}\ , \quad \bar{\gamma}_{.} = \frac{1}{mr}\sum_{j,l}^{m,r} \gamma_j^{(l)}\ , \quad \bar{\rho} = \frac{1}{r}\sum_{l=1}^{r} \rho_{(l)}\ ,$$

we have

$$SSA = m\sum_{l=1}^{r}\sum_{i=1}^{m}(\bar{Y}_{i..}^{(l)} - \bar{Y}_{...}^{(l)})^2 = m\sum_{l=1}^{r}\sum_{i=1}^{m}(\alpha_i^{(l)} - \bar{\alpha}_{.}^{(l)} + \bar{\varepsilon}_{i..}^{(l)} - \bar{\varepsilon}_{...}^{(l)})^2,$$

$$SSC = m\sum_{l=1}^{r}\sum_{j=1}^{m}(\bar{Y}_{.j.}^{(l)} - \bar{Y}_{...}^{(l)})^2 = m\sum_{l=1}^{r}\sum_{j=1}^{m}(\gamma_j^{(l)} - \bar{\gamma}_{.}^{(l)} + \bar{\varepsilon}_{.j.}^{(l)} - \bar{\varepsilon}_{...}^{(l)})^2,$$

$$SST = mr\sum_{k=1}^{m}(\bar{Y}_{..k} - \bar{Y}_{...})^2 = mr\sum_{k=1}^{m}(\tau_k - \bar{\tau}_{.} + \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})^2,$$

$$SSP = m^2\sum_{l=1}^{r}(\bar{Y}_{...}^{(l)} - \bar{Y}_{...})^2 = m^2\sum_{l=1}^{r}(\bar{\alpha}_{.}^{(l)} - \bar{\alpha}_{.} + \bar{\gamma}_{.}^{(l)} - \bar{\gamma}_{.} + \rho_{(l)} - \bar{\rho} + \bar{\varepsilon}_{...}^{(l)} - \bar{\varepsilon}_{...})^2,$$

$$SSE = \sum_{l=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{m}(Y_{ijk}^{(l)} - \bar{Y}_{i..}^{(l)} - \bar{Y}_{.j.}^{(l)} - \bar{Y}_{..k} + \bar{Y}_{...}^{(l)} + \bar{Y}_{...})^2$$

$$= \sum_{l=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{m}(\varepsilon_{ijk}^{(l)} - \bar{\varepsilon}_{i..}^{(l)} - \bar{\varepsilon}_{.j.}^{(l)} - \bar{\varepsilon}_{..k} + \bar{\varepsilon}_{...}^{(l)} + \bar{\varepsilon}_{...})^2 .$$

To evaluate the expected values of these sums of squares by Theorem 3.2, we shall need the variances and covariances of the following sets of variables,

$$\alpha_i^{(l)} + \bar{\varepsilon}_{i..}^{(l)}\ , \quad \gamma_j^{(l)} + \bar{\varepsilon}_{.j.}^{(l)}\ , \quad \tau_k + \bar{\varepsilon}_{..k}\ ,$$

$$\bar{\alpha}_{.}^{(l)} + \bar{\gamma}_{.}^{(l)} + \rho_{(l)} + \bar{\varepsilon}_{...}^{(l)} .$$

Therefore we must compute

$$E[(\alpha_i^{(\ell)})^2]\ ,\ E(\alpha_i^{(\ell)}\alpha_{i'}^{(\ell)})\ ,\ E[(\gamma_j^{(\ell)})^2]\ ,\ E(\gamma_j^{(\ell)}\gamma_{j'}^{(\ell)})\ ,$$
$$E(\tau_k^2)\ ,\ E(\tau_k\tau_{k'})\ ,\ E[(\bar{\alpha}_.^{(\ell)})^2],\ E(\bar{\alpha}_.^{(\ell)}\bar{\alpha}_.^{(\ell')}),\ E[(\bar{\gamma}_.^{(\ell)})^2],$$
$$E(\bar{\gamma}_.^{(\ell)}\bar{\gamma}_.^{(\ell')})\ ,\ E(\rho_{(\ell)}^2)\ ,\ E(\rho_{(\ell)}\rho_{(\ell')})$$

in a form which will be valid regardless of the nature of the populations used.

If $\alpha_i^{(\ell)}$ comes from a Type II population, $E[(\alpha_i^{(\ell)})^2] = \sigma^2_{\alpha(\ell)}$, and if it comes from a Type III population,

$$E[(\alpha_i^{(\ell)})^2] = \frac{1}{M_{\alpha(\ell)}}\sum_{i=1}^{M_{\alpha(\ell)}}(\alpha_i^{(\ell)})^2 = \left(1-\frac{1}{M_{\alpha(\ell)}}\right)\sigma^2_{\alpha(\ell)}\ ,$$

which gives the formula for the Type II case if we let $M_{\alpha(\ell)}\to\infty$. For the Type II case, $E(\alpha_i^{(\ell)}\alpha_{i'}^{(\ell)}) = 0$, and for the Type III case

$$E(\alpha_i^{(\ell)}\alpha_{i'}^{(\ell)}) = \sum_{i\neq i'}^{M_{\alpha(\ell)}}\frac{\alpha_i^{(\ell)}\alpha_{i'}^{(\ell)}}{M_{\alpha(\ell)}(M_{\alpha(\ell)}-1)} = -\sum_{i=1}^{M_{\alpha(\ell)}}\frac{(\alpha_i^{(\ell)})^2}{M_{\alpha(\ell)}(M_{\alpha(\ell)}-1)} = -\frac{\sigma^2_{\alpha(\ell)}}{M_{\alpha(\ell)}}\ ,$$

since

$$0 = \sum_{i,i'}^{M_{\alpha(\ell)}}\alpha_i^{(\ell)}\alpha_{i'}^{(\ell)} = \sum_{i=1}^{M_{\alpha(\ell)}}(\alpha_i^{(\ell)})^2 + \sum_{i\neq i'}^{M_{\alpha(\ell)}}\alpha_i^{(\ell)}\alpha_{i'}^{(\ell)}\ .$$

Thus the formula for the Type II case is again included in the Type III case. Similarly for the $\gamma_j^{(\ell)}$'s , τ_k's and $\rho_{(\ell)}$'s ,

$$E(\gamma_j^{(\ell)})^2 = \left(1-\frac{1}{M_{\gamma(\ell)}}\right)\sigma^2_{\gamma(\ell)}\ ,\ E(\gamma_j^{(\ell)}\gamma_{j'}^{(\ell)}) = -\frac{\sigma^2_{\gamma(\ell)}}{M_{\gamma(\ell)}}\ ,$$

$$E(\gamma_k^2) = \left(1 - \frac{1}{M_\gamma}\right)\sigma_\gamma^2 \quad , \quad E(\gamma_k \gamma_{k'}) = -\frac{\sigma_\gamma^2}{M_\gamma} \quad ,$$

and

$$E(\rho_{(\ell)}^2) = \left(1 - \frac{1}{M_\rho}\right)\sigma_\rho^2 \quad , \quad E(\rho_{(\ell)}\rho_{(\ell')}) = -\frac{\sigma_\rho^2}{M_\rho} \quad .$$

Also for a Type III population,

$$E[(\bar{\alpha}_.^{(\ell)})^2] = \frac{1}{m^2}\sum_{i,i'}^{m} E(\alpha_i^{(\ell)}\alpha_{i'}^{(\ell)})$$

$$= \frac{1}{m^2}\sum_{i=1}^{m} E[(\alpha_i^{(\ell)})^2] + \frac{1}{m^2}\sum_{i \neq i'}^{m} E(\alpha_i^{(\ell)}\alpha_{i'}^{(\ell)})$$

$$= \frac{1}{m}\left(1 - \frac{1}{M_{\alpha^{(\ell)}}}\right)\sigma_{\alpha(\ell)}^2 - \frac{(m^2 - m)}{m^2}\,\frac{\sigma_{\alpha(\ell)}^2}{M_{\alpha^{(\ell)}}}$$

$$= \left(\frac{1}{m} - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 \quad ,$$

and

$$E(\bar{\alpha}_.^{(\ell)}\bar{\alpha}_.^{(\ell')}) = 0 \quad .$$

Similarly,

$$E[(\bar{\gamma}_.^{(\ell)})^2] = \left(\frac{1}{m} - \frac{1}{M_{\gamma(\ell)}}\right)\sigma_{\gamma(\ell)}^2 \quad ,$$

$$E(\bar{\gamma}_.^{(\ell)}\bar{\gamma}_.^{(\ell')}) = 0 \quad .$$

To evaluate E(SSA) we let

$$y_i = \alpha_i^{(\ell)} + \bar{\varepsilon}_{i..}^{(\ell)} \quad ,$$

$$\bar{y}_{\cdot} = \frac{1}{m}\sum_{i=1}^{m} y_i = \bar{\alpha}_{\cdot}^{(\ell)} + \bar{\varepsilon}_{\cdots}^{(\ell)} ,$$

so that

$$SSA = m \sum_{\ell=1}^{\Lambda} \sum_{i=1}^{m} (y_i - \bar{y}_{\cdot})^2 ,$$

$$\mu_i = E(y_i) = (1-\delta_\alpha)\,\alpha_i^{(\ell)} , \quad \bar{\mu}_{\cdot} = \frac{1}{m}\sum_{i=1}^{m} \mu_i = 0 ,$$

where $\delta_\alpha = 0$ if the $\alpha_i^{(\ell)}$'s come from a Type I population and $\delta_\alpha = 1$ otherwise. Also

$$\begin{aligned} \sigma_i^2 = Var(y_i) &= E\left[\alpha_i^{(\ell)} + \bar{\varepsilon}_{i\cdot\cdot}^{(\ell)} - E(\alpha_i^{(\ell)})\right]^2 \\ &= E\left[\alpha_i^{(\ell)} - E(\alpha_i^{(\ell)})\right]^2 + E\left[(\bar{\varepsilon}_{i\cdot\cdot}^{(\ell)})^2\right] \\ &= \delta_\alpha \left(1 - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 + \frac{\sigma^2}{m} , \end{aligned}$$

and

$$\bar{\sigma}_{\cdot}^2 = \frac{1}{m}\sum_{i=1}^{m} \sigma_i^2 = \delta_\alpha \left(1 - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 + \frac{\sigma^2}{m} .$$

Next

$$\begin{aligned} \lambda = cov(y_i, y_{i'}) &= E\left\{\left[\alpha_i^{(\ell)} - E(\alpha_i^{(\ell)})\right]\left[\alpha_{i'}^{(\ell)} - E(\alpha_{i'}^{(\ell)})\right]\right\} \\ &= -\frac{\delta_\alpha\,\sigma_{\alpha(\ell)}^2}{M_{\alpha(\ell)}} , \end{aligned}$$

$$\bar{\sigma}_{\cdot}^2 - \lambda = \delta_\alpha\,\sigma_{\alpha(\ell)}^2 + \frac{\sigma^2}{m} ,$$

$$E(SSA) = (1-\delta_\alpha)\, m \sum_{l=1}^{r} \sum_{i=1}^{m} (\alpha_i^{(l)})^2 + m(m-1) \sum_{l=1}^{r} \left(\delta_\alpha \sigma_{\alpha(l)}^2 + \frac{\sigma^2}{m}\right)$$

$$= r(m-1)\sigma^2 + \delta_\alpha m(m-1) \sum_{l=1}^{r} \sigma_{\alpha(l)}^2 + (1-\delta_\alpha)\, m \sum_{l=1}^{r} \sum_{i=1}^{m} (\alpha_i^{(l)})^2$$

and

$$E(MSA) = \sigma^2 + \delta_\alpha \frac{m}{r} \sum_{l=1}^{r} \sigma_{\alpha(l)}^2 + (1-\delta_\alpha) \frac{m}{r(m-1)} \sum_{l=1}^{r} \sum_{i=1}^{m} (\alpha_i^{(l)})^2 .$$

Similarly,

$$E(MSC) = \sigma^2 + \delta_\gamma \frac{m}{r} \sum_{l=1}^{r} \sigma_{\gamma(l)}^2 + (1-\delta_\gamma) \frac{m}{r(m-1)} \sum_{l=1}^{r} \sum_{j=1}^{m} (\gamma_j^{(l)})^2 .$$

To evaluate E(SST) we let

$$y_k = \tau_k + \bar{\varepsilon}_{..k} \quad , \quad \bar{y}_. = \frac{1}{m} \sum_{k=1}^{m} y_k = \bar{\tau}_. + \bar{\varepsilon}_{...} \quad ,$$

so that

$$SST = mr \sum_{k=1}^{m} (y_k - \bar{y}_.)^2 ,$$

$$\mu_k = E(y_k) = (1-\delta_\tau)\tau_k \quad , \quad \bar{\mu}_. = \frac{1}{m} \sum_{k=1}^{m} \mu_k = 0 \quad ,$$

where $\delta_\tau = 0$ if the τ_k's come from a Type I population and $\delta_\tau = 1$ otherwise. Also

$$\sigma_k^2 = Var(y_k) = E[\tau_k - E(\tau_k) + \bar{\varepsilon}_{..k}]^2$$

$$= E[\tau_k - E(\tau_k)]^2 + E(\bar{\varepsilon}_{..k}^2)$$

$$= \delta_\gamma\left(1 - \frac{1}{M_\gamma}\right)\sigma_\gamma^2 + \frac{\sigma^2}{mr} ,$$

and

$$\overline{\sigma_.}^2 = \frac{1}{m}\sum_{k=1}^{m} \sigma_k^2 = \delta_\gamma\left(1 - \frac{1}{M_\gamma}\right)\sigma_\gamma^2 + \frac{\sigma^2}{mr} .$$

Next

$$\lambda = \operatorname{cov}(y_k, y_{k'}) = E\{[\gamma_k - E(\gamma_k)][\gamma_{k'} - E(\gamma_{k'})]\}$$

$$= -\delta_\gamma \frac{\sigma_\gamma^2}{M_\gamma} ,$$

$$\overline{\sigma_.}^2 - \lambda = \delta_\gamma \sigma_\gamma^2 + \frac{\sigma^2}{mr} ,$$

$$E(SST) = (1-\delta_\gamma)\, mr \sum_{k=1}^{m} \gamma_k^2 + r\, m(m-1)\left(\delta_\gamma \sigma_\gamma^2 + \frac{\sigma^2}{mr}\right)$$

$$= (m-1)\sigma^2 + \delta_\gamma\, r\, m(m-1)\sigma_\gamma^2 + (1-\delta_\gamma)\, mr \sum_{k=1}^{m} \gamma_k^2$$

and

$$E(MST) = \sigma^2 + \delta_\gamma\, mr\, \sigma_\gamma^2 + (1-\delta_\gamma)\frac{mr}{m-1}\sum_{k=1}^{m} \gamma_k^2 .$$

To evaluate E(SSP) we let

$$y_\ell = \bar{\alpha}_.^{(\ell)} + \bar{\gamma}_.^{(\ell)} + \rho_{(\ell)} + \bar{\varepsilon}_{\dots}^{(\ell)} ,$$

$$\bar{y}_. = \frac{1}{r}\sum_{\ell=1}^{r} y_\ell = \bar{\alpha}_. + \bar{\gamma}_. + \bar{\rho} + \bar{\varepsilon}_{\dots} ,$$

so that

$$SSP = m^2 \sum_{\ell=1}^{r} (y_\ell - \bar{y}_.)^2 ,$$

$$\mu_\ell = E(y_\ell) = (1-\delta_\rho)(\bar{\alpha}_.^{(\ell)} + \bar{\gamma}_.^{(\ell)} + \rho_{(\ell)}) = (1-\delta_\rho)\rho_{(\ell)}$$

$$\bar{\mu}_. = \frac{1}{r}\sum_{\ell=1}^{r}\mu_\ell = 0 \quad ,$$

where $\delta_\rho = 0$ if the $\rho_{(\ell)}$'s come from a Type I population and $\delta_\rho = 1$ otherwise. Also

$$\sigma_\ell^2 = Var(y_\ell) = E\left[\bar{\alpha}_.^{(\ell)} - E(\bar{\alpha}_.^{(\ell)}) + \bar{\gamma}_.^{(\ell)} - E(\bar{\gamma}_.^{(\ell)}) + \rho_{(\ell)} - E(\rho_{(\ell)}) + \bar{\varepsilon}_{...}^{(\ell)}\right]^2$$

$$= E[\bar{\alpha}_.^{(\ell)} - E(\bar{\alpha}_.^{(\ell)})]^2 + E[\bar{\gamma}_.^{(\ell)} - E(\bar{\gamma}_.^{(\ell)})]^2 + E[\rho_{(\ell)} - E(\rho_{(\ell)})]^2 + E[(\bar{\varepsilon}_{...}^{(\ell)})^2]$$

$$= \left(\frac{1}{m} - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 + \left(\frac{1}{m} - \frac{1}{M_{\gamma(\ell)}}\right)\sigma_{\gamma(\ell)}^2$$

$$+ \delta_\rho\left(1 - \frac{1}{M_\rho}\right)\sigma_\rho^2 + \frac{\sigma^2}{m^2} \quad ,$$

since, for the Type I model, $M_{\alpha(\ell)} = M_{\gamma(\ell)} = m$, and

$$\bar{\sigma}_.^2 = \frac{1}{r}\sum_{\ell=1}^{r}\sigma_\ell^2 = \frac{1}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 + \frac{1}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\gamma(\ell)}}\right)\sigma_{\gamma(\ell)}^2$$

$$+ \delta_\rho\left(1 - \frac{1}{M_\rho}\right)\sigma_\rho^2 + \frac{\sigma^2}{m^2} \quad .$$

Next

$$\lambda = cov(y_\ell, y_{\ell'}) = E\Big\{\left[\bar{\alpha}_.^{(\ell)} - E(\bar{\alpha}_.^{(\ell)}) + \bar{\gamma}_.^{(\ell)} - E(\bar{\gamma}_.^{(\ell)}) + \rho_{(\ell)} - E(\rho_{(\ell)})\right]\Big[$$

$$\bar{\alpha}_.^{(\ell')} - E(\bar{\alpha}_.^{(\ell')}) + \bar{\gamma}_.^{(\ell')} - E(\bar{\gamma}_.^{(\ell')} + \rho_{(\ell')} - E(\rho_{(\ell')})\Big]\Big\}$$

$$= E\{[\rho_{(\ell)} - E(\rho_{(\ell)})][\rho_{(\ell')} - E(\rho_{(\ell')})]\}$$

$$= -\delta_\rho \frac{\sigma_\rho^2}{M_\rho} ,$$

$$\bar{\sigma}_.^2 - \lambda = \frac{1}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 + \frac{1}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\gamma(\ell)}}\right)\sigma_{\gamma(\ell)}^2$$

$$+ \delta_\rho \sigma_\rho^2 + \frac{\sigma^2}{m^2} ,$$

$$E(SSP) = (1-\delta_\rho) m^2 \sum_{\ell=1}^{r} \rho_{(\ell)}^2 + m^2(r-1)\Big[\frac{1}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 +$$

$$\frac{1}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\gamma(\ell)}}\right)\sigma_{\gamma(\ell)}^2 + \delta_\rho \sigma_\rho^2 + \frac{\sigma^2}{m^2}\Big]$$

$$= (r-1)\sigma^2 + \delta_\rho m^2(r-1)\sigma_\rho^2 + \frac{m^2(r-1)}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2$$

$$+ \frac{m^2(r-1)}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\gamma(\ell)}}\right)\sigma_{\gamma(\ell)}^2 + (1-\delta_\rho) m^2 \sum_{\ell=1}^{r} \rho_{(\ell)}^2$$

and

$$E(MSP) = \sigma^2 + \frac{m^2}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 + \frac{m^2}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\gamma(\ell)}}\right)\sigma_{\gamma(\ell)}^2$$

$$+ \delta_\rho m^2 \sigma_\rho^2 + (1-\delta_\rho)\frac{m^2}{r-1}\sum_{\ell=1}^{r} \rho_{(\ell)}^2 .$$

Finally, by the theory of Chapter II, we know that SSE is distributed as $\chi^2\sigma^2$ with (m-1)(mr-r-1) d.f. . Hence

$$E\left(\frac{SSE}{\sigma^2}\right) = (m-1)(mr-r-1)$$

and

$$E(MSE) = \sigma^2 \quad .$$

If we define

$$\sigma^2_{\alpha(\ell)} = \frac{1}{m-1}\sum_{i=1}^{m}(\alpha_i^{(\ell)})^2 \quad , \quad \sigma^2_{\gamma(\ell)} = \frac{1}{m-1}\sum_{j=1}^{m}(\gamma_j^{(\ell)})^2 \ ,$$

$$\sigma_\tau^2 = \frac{1}{m-1}\sum_{K=1}^{m}\tau_K^2 \quad , \quad \sigma_\rho^2 = \frac{1}{r-1}\sum_{\ell=1}^{r}\rho_{(\ell)}^2$$

for the Type I populations, the above results can be summarized in table 6.4.

The following formulas are more convenient for computation:

$$SSA = m\sum_{\ell=1}^{r}\sum_{i=1}^{m}(\bar{Y}_{i..}^{(\ell)} - \bar{Y}_{...}^{(\ell)})^2 = \sum_{\ell=1}^{r}\left\{\frac{(\bar{Y}_{i..}^{(\ell)})^2}{m} - \frac{(\bar{Y}_{...}^{(\ell)})^2}{m^2}\right.$$

$$= \sum_{\ell=1}^{r}\frac{(Y_{i..}^{(\ell)})^2}{m} - \sum_{\ell=1}^{r}\frac{(Y_{...}^{(\ell)})^2}{m^2} \quad ,$$

$$SSC = m\sum_{\ell=1}^{r}\sum_{j=1}^{m}(\bar{Y}_{.j.}^{(\ell)} - \bar{Y}_{...}^{(\ell)})^2 = \sum_{\ell=1}^{r}\frac{(Y_{.j.}^{(\ell)})^2}{m} - \sum_{\ell=1}^{r}\frac{(Y_{...}^{(\ell)})^2}{m^2} \quad ,$$

$$SST = mr\sum_{K=1}^{m}(\bar{Y}_{..K} - \bar{Y}_{...})^2 = \sum_{K=1}^{m}\frac{Y_{..K}^2}{mr} - \frac{Y_{...}^2}{m^2 r} \quad ,$$

$$SSP = m^2\sum_{\ell=1}^{r}(\bar{Y}_{...}^{(\ell)} - \bar{Y}_{...})^2 = \sum_{\ell=1}^{r}\frac{Y_{...}^{(\ell)2}}{m^2} - \frac{Y_{...}^2}{m^2 r} \quad ,$$

$$\sum_{\ell=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{m}(Y_{ijK}^{(\ell)} - \bar{Y}_{...})^2 = \sum_{\ell=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{m}(Y_{ijK}^{(\ell)})^2 - \frac{Y_{...}^2}{m^2 r} \quad ,$$

and SSE is obtained by subtraction.

TABLE 6.4

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Rows	r(m-1)	$SSA = m\sum_{\ell=1}^{r}\sum_{i=1}^{m}(\bar{Y}_{i..}^{(\ell)} - \bar{Y}_{...}^{(\ell)})^2$	$MSA = \frac{SSA}{r(m-1)}$	$\sigma^2 + \frac{m}{r}\sum_{\ell=1}^{r}\sigma_{\alpha(\ell)}^2$
Columns	r(m-1)	$SSC = m\sum_{\ell=1}^{r}\sum_{j=1}^{m}(\bar{Y}_{.j.}^{(\ell)} - \bar{Y}_{...}^{(\ell)})^2$	$MSC = \frac{SSC}{r(m-1)}$	$\sigma^2 + \frac{m}{r}\sum_{\ell=1}^{r}\sigma_{\gamma(\ell)}^2$
Treatments	m-1	$SST = mr\sum_{k=1}^{m}(\bar{Y}_{..k} - \bar{Y}_{...})^2$	$MST = \frac{SST}{m-1}$	$\sigma^2 + mr\sigma_{\tau}^2$
Replications	r-1	$SSP = m^2\sum_{\ell=1}^{r}(\bar{Y}_{...}^{(\ell)} - \bar{Y}_{...})^2$	$MSP = \frac{SSP}{r-1}$	$\sigma^2 + m^2\sigma_{\rho}^2 + \frac{m^2}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\alpha(\ell)}}\right)\sigma_{\alpha(\ell)}^2 + \frac{m^2}{r}\sum_{\ell=1}^{r}\left(\frac{1}{m} - \frac{1}{M_{\gamma(\ell)}}\right)\sigma_{\gamma(\ell)}^2$
Error	(m-1)(mr-r-1)	$SSE = \sum_{\ell=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{m}(Y_{ijk}^{(\ell)} - \bar{Y}_{i..}^{(\ell)} - \bar{Y}_{.j.}^{(\ell)} - \bar{Y}_{..k} + \bar{Y}_{...}^{(\ell)} + \bar{Y}_{...})^2$	$MSE = \frac{SSE}{(m-1)(mr-r-1)}$	σ^2
Total	m^2r-1	$\sum_{\ell=1}^{r}\sum_{i=1}^{m}\sum_{j=1}^{m}(Y_{ijk}^{(\ell)} - \bar{Y}_{...})^2$		

6.8 Distributions of the Sums of Squares for Replicated Latin Squares

Corresponding to the hypotheses

H_1: $\alpha_i^{(\ell)} = 0$, H_2: $\gamma_j^{(\ell)} = 0$, H_3: $\tau_k = 0$, H_4: $\rho_{(\ell)} = 0$,

we have the hypotheses

$$\sigma_{\alpha(\ell)} = 0 \ , \quad \sigma_{\gamma(\ell)} = 0 \ , \quad \sigma_{\tau} = 0 \ , \quad \sigma_{\rho} = 0 \ ,$$

if the corresponding variables are from other than a Type I population. Then, since the populations have zero means, it follows that the corresponding variables are equal to zero.

In the case of the Type I model, the theory of Chapter II implies that

$$\bar{Y}_{i..}^{(\ell)} - \bar{Y}_{...}^{(\ell)} , \quad \bar{Y}_{.j.}^{(\ell)} - \bar{Y}_{...}^{(\ell)} , \quad \bar{Y}_{..k} - \bar{Y}_{...} , \quad \bar{Y}_{...}^{(\ell)} - \bar{Y}_{...} ,$$

$$(i,j,k = 1,2,\ldots,m-1; \ell = 1,2,\ldots,r)$$

are distributed independently of SSE. Therefore any function of these expressions is distributed independently of SSE and, in particular, this holds for

$$\bar{Y}_{m..}^{(\ell)} - \bar{Y}_{...}^{(\ell)} , \quad \bar{Y}_{.m.}^{(\ell)} - \bar{Y}_{...}^{(\ell)} , \quad \bar{Y}_{..m} - \bar{Y}_{...} , \quad \bar{Y}_{...}^{(\ell)} - \bar{Y}_{...} ,$$

$(\ell = 1,2,\ldots,r)$. These results were obtained for the model

$$Y_{ijk}^{(\ell)} = \mu + \alpha_i^{(\ell)} + \gamma_j^{(\ell)} + \tau_k + \rho_{(\ell)} + \varepsilon_{ijk}^{(\ell)}$$

and they hold for the particular case where $Y_{ijk}^{(\ell)} = \varepsilon_{ijk}^{(\ell)}$.

Hence

$$\bar{\varepsilon}_{i..}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)} , \quad \bar{\varepsilon}_{.j.}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)} , \quad \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...} , \quad \bar{\varepsilon}_{...}^{(\ell)} - \bar{\varepsilon}_{...} ,$$

are distributed independently of SSE.

We shall now show that these expressions in the $\varepsilon_{ijk}^{(\ell)}$'s are independent of each other. Since they have normal distributions, it is sufficient to prove that

$$\operatorname{cov}(\bar{\varepsilon}_{i..}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)}, \bar{\varepsilon}_{.j.}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)}) = \operatorname{cov}(\bar{\varepsilon}_{i..}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)}, \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})$$

$$= \operatorname{cov}(\bar{\varepsilon}_{i..}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)}, \bar{\varepsilon}_{...}^{(\ell)} - \bar{\varepsilon}_{...}) = \operatorname{cov}(\bar{\varepsilon}_{.j.}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)}, \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})$$

$$= \operatorname{cov}(\bar{\varepsilon}_{.j.}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)}, \bar{\varepsilon}_{...}^{(\ell)} - \bar{\varepsilon}_{...}) = \operatorname{cov}(\bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...}, \bar{\varepsilon}_{...}^{(\ell)} - \bar{\varepsilon}_{...}) = 0 .$$

Since

$$E(\bar{\varepsilon}_{i..}^{(\ell)} \bar{\varepsilon}_{.j.}^{(\ell)}) = \sigma^2/m^2 , \quad E(\bar{\varepsilon}_{i..}^{(\ell)} \bar{\varepsilon}_{...}^{(\ell)}) = \frac{m}{m^3}\sigma^2 = \sigma^2/m^2 ,$$

$$E(\bar{\varepsilon}_{.j.}^{(\ell)} \bar{\varepsilon}_{...}^{(\ell)}) = \frac{m}{m^3}\sigma^2 = \sigma^2/m^2 , \quad E[(\bar{\varepsilon}_{...}^{(\ell)})^2] = \sigma^2/m^2 ,$$

$$E(\bar{\varepsilon}_{i..}^{(\ell)} \bar{\varepsilon}_{..k}) = \frac{\sigma^2}{m^2 r} ,$$

$$E(\bar{\varepsilon}_{i..}^{(\ell)} \bar{\varepsilon}_{...}) = \frac{m}{m^3 r}\sigma^2 = \frac{\sigma^2}{m^2 r} ,$$

$$E(\bar{\varepsilon}_{...}^{(\ell)} \bar{\varepsilon}_{..k}) = \frac{m}{m^3 r}\sigma^2 = \frac{\sigma^2}{m^2 r} ,$$

$$E(\bar{\varepsilon}_{...}^{(\ell)} \bar{\varepsilon}_{...}) = \frac{m^2}{m^4 r}\sigma^2 = \frac{\sigma^2}{m^2 r} ,$$

$$E(\bar{\varepsilon}_{.j.}^{(\ell)} \bar{\varepsilon}_{..k}) = \frac{\sigma^2}{m^2 r}$$

$$E(\bar{\varepsilon}_{.j.}^{(\ell)} \bar{\varepsilon}_{...}) = \frac{m}{m^3 r}\sigma^2 = \frac{\sigma^2}{m^2 r} ,$$

$$E(\bar{\varepsilon}_{..k} \bar{\varepsilon}_{...}) = \frac{mr}{m^3 r^2}\sigma^2 = \frac{\sigma^2}{m^2 r} ,$$

it then follows that all the above covariances are equal to zero.

In section 6.5 we tested the hypotheses H_1, H_2, H_3, and H_4 for the Type I model by the statistics

$$F_1 = \frac{MSA}{MSE}, \quad F_2 = \frac{MSC}{MSE}, \quad F_3 = \frac{MST}{MSE}, \quad F_4 = \frac{MSP}{MSE},$$

respectively, since subject to the corresponding hypotheses,

$$SSA = m \sum_{\ell=1}^{r} \sum_{i=1}^{m} (\bar{\varepsilon}_{i..}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)})^2, \quad SSC = m \sum_{\ell=1}^{r} \sum_{j=1}^{m} (\bar{\varepsilon}_{.j.}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)})^2,$$

$$SST = mr \sum_{k=1}^{m} (\bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})^2, \quad SSP = m^2 \sum_{\ell=1}^{r} (\bar{\varepsilon}_{...}^{(\ell)} - \bar{\varepsilon}_{...})^2,$$

and

$$SSE = \sum_{\ell=1}^{r} \sum_{i=1}^{m} \sum_{j=1}^{m} (\varepsilon_{ijk}^{(\ell)} - \bar{\varepsilon}_{i..}^{(\ell)} - \bar{\varepsilon}_{.j.}^{(\ell)} - \bar{\varepsilon}_{..k} + \bar{\varepsilon}_{...}^{(\ell)} + \bar{\varepsilon}_{...})^2$$

are independently distributed as $\chi^2\sigma^2$, with r(m-1), r(m-1), m-1, r-1, and (m-1)(mr-r-1) d.f. respectively.

We shall now determine what tests can be made when we are not dealing with a Type I model. In section 6.7 we saw that

$$SSA = m \sum_{\ell=1}^{r} \sum_{i=1}^{m} (\alpha_i^{(\ell)} - \bar{\alpha}_.^{(\ell)} + \bar{\varepsilon}_{i..}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)})^2,$$

$$SSC = m \sum_{\ell=1}^{r} \sum_{j=1}^{m} (\gamma_j^{(\ell)} - \bar{\gamma}_.^{(\ell)} + \bar{\varepsilon}_{.j.}^{(\ell)} - \bar{\varepsilon}_{...}^{(\ell)})^2,$$

$$SST = mr \sum_{k=1}^{m} (\tau_k - \bar{\tau}_. + \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})^2,$$

and

$$SSP = m^2 \sum_{\ell=1}^{r} (\bar{\alpha}_.^{(\ell)} - \bar{\alpha}_. + \bar{\gamma}_.^{(\ell)} - \bar{\gamma}_. + \rho_{(\ell)} - \bar{\rho} + \bar{\varepsilon}_{...}^{(\ell)} - \bar{\varepsilon}_{...})^2.$$

Subject to H_1, H_2, and H_3, the sums of squares SSA, SSC, and SST reduce to the corresponding expressions for the Type I model and the same tests apply. However it is impossible to test H_4 for the Type II or Type III models unless some approximation device is employed. In fact, for the Type III model we can never hope to carry out an F test for H_4 since such a test requires the $\alpha_i^{(\ell)}$'s and $\gamma_j^{(\ell)}$'s to be normally distributed.

6.9 The Type I Model for Orthogonal Latin Squares

We shall now extend the methods of the previous sections to the case where we have r orthogonal $m \times m$ Latin squares where $r \leq m-1$. Denote the observations by $Y_{ijk_1,k_2,\ldots,k_r}$, where $i,j = 1,2,\ldots,m$ represent the row and column numbers respectively, k_s $(s = 1,2,\ldots,r)$, takes on the values from 1 to m and represents the m treatments of the sth Latin square. Our model is given by

$$Y_{ijk_1,\ldots,k_r} = \mu + \alpha_i + \gamma_j + \tau_{k_1}^{(1)} + \tau_{k_2}^{(2)} + \cdots + \tau_{k_r}^{(r)} + \varepsilon_{ijk_1,\ldots,k_r}$$

where α_i, γ_j, $\tau_{k_\ell}^{(\ell)}$ $(\ell = 1,2,\ldots,r)$, represent the row, column, and treatment effects respectively. The parameters are subject to the restrictions

$$\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{m} \gamma_j = \sum_{k_1=1}^{m} \tau_{k_1}^{(1)} = \cdots = \sum_{k_r=1}^{m} \tau_{k_r}^{(r)} = 0 .$$

The subscripts k_ℓ, $\ell = 1,2,\ldots,r$ are functions of i and j such that for a fixed i (j) they take on each of the values from 1 to m exactly once in some order as j (i) takes on the values from 1 to m. The pair of numbers (k_ℓ, k_m) takes on every possible ordered pair of numbers exactly once where k_ℓ and k_m are selected independently from the numbers 1 to m . From Bedrosian's thesis we exhibit the results for this model in table 6.5, where $\overline{Y}_i$, $\overline{Y}_j$ are the means of the observations of the ith row and jth column respectively; $\overline{Y}_{k_\ell}^{(\ell)}$ $(\ell = 1,2,\ldots,r)$ is the mean of the m values of $Y_{ijk_1,\ldots,k_r}$ which have the given number k_ℓ as the $(\ell + 2)$nd subscript and $\overline{Y}$ is the grand mean.

We shall be interested in testing the hypotheses

H_1: $\alpha_i = 0$ $(i = 1,2,\ldots,m)$,

H_2: $\gamma_j = 0$ $(j = 1,2,\ldots,m)$,

H_3: $\tau_{k_1}^{(1)} = 0$ $(k_1 = 1,2,\ldots,m)$,

H_4: $\tau_{k_2}^{(2)} = 0$ $(k_2 = 1,2,\ldots,m)$,

. . . . ,

H_{r+2}: $\tau_{k_r}^{(r)} = 0$ $(k_r = 1,2,\ldots,m)$.

The theory of Chapter II tells us that, subject to the corresponding hypotheses, SSA, SSC, $SST_1,\ldots$, SST_r, and SSE are independently distributed as $\chi^2\sigma^2$, the first r+2 sums with m-1 d.f. and the last sum with (m-1)(m-r-1) d.f. . The hypotheses H_1, $H_2,\ldots$, H_{r+2} are tested by the statistics

$$F_1 = \frac{MSA}{MSE}, \quad F_2 = \frac{MSC}{MSE}, \quad F_3 = \frac{MST_1}{MSE}, \ldots, \quad F_{r+2} = \frac{MST_r}{MSE},$$

respectively.

6.10 Other Models for Orthogonal Latin Squares

Our regression model is

$$Y_{ijk_1,\ldots,k_r} = \mu + \alpha_i + \gamma_j + \tau_{k_1}^{(1)} + \tau_{k_2}^{(2)} + \cdots + \tau_{k_r}^{(r)} + \varepsilon_{ijk_1,\ldots,k_r}$$

For the Type II model we assume that the α_i's , γ_j's , $\tau_{k_\ell}^{(\ell)}$'s $(\ell = 1,2,\ldots,r)$, and $\varepsilon_{ijk_1,\ldots,k_r}$ are NID with zero means and variances σ_α^2, σ_γ^2, $\sigma_{\tau_\ell}^2$ $(\ell = 1,2,\ldots,r)$, and σ^2 respec-

TABLE 6.5

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Rows	m-1	$SSA = m\sum_{i=1}^{m}(\bar{Y}_{i\cdot} - \bar{Y})^2$	$MSA = \frac{SSA}{m-1}$
Columns	m-1	$SSC = m\sum_{j=1}^{m}(\bar{Y}_{\cdot j} - \bar{Y})^2$	$MSC = \frac{SSC}{m-1}$
Treatments	m-1	$SST_1 = m\sum_{k_1=1}^{m}(\bar{Y}^{(1)}_{k_1} - \bar{Y})^2$	$MST_1 = \frac{SST_1}{m-1}$
	...	....	....
	m-1	$SST_r = m\sum_{k_r=1}^{m}(\bar{Y}^{(r)}_{k_r} - \bar{Y})^2$	$MST_r = \frac{SST_r}{m-1}$
Error	(m-1)(m-r-1)	$SSE = \sum_{i=1}^{m}\sum_{j=1}^{m}[Y_{ijk_1,\ldots,k_r} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} - \bar{Y}^{(1)}_{k_1} - \bar{Y}^{(2)}_{k_2} - \cdots - \bar{Y}^{(r)}_{k_r} + (r+1)\bar{Y}]^2$	$MSE = \frac{SSE}{(m-1)(m-r-1)}$
Total	m^2-1	$\sum_{i=1}^{m}\sum_{j=1}^{m}(Y_{ijk_1,\ldots,k_r} - \bar{Y})^2$	

tively. Then we have

$$E(Y_{ijk_1,\ldots,k_r}) = \mu$$

and

$$Var(Y_{ijk_1,\ldots,k_r}) = \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_{\tau_1}^2 + \sigma_{\tau_2}^2 + \cdots + \sigma_{\tau_r}^2 + \sigma^2 .$$

For the Type III model we assume that the α_i's , γ_j's , $\tau_{k_\ell}^{(\ell)}$'s ($\ell = 1,2,\ldots,r$), come from finite independent populations of size M_α, M_γ, M_{τ_ℓ} ($\ell = 1,2,\ldots,r$), say, respectively, with zero means and variances

$$\sigma_\alpha^2 = \frac{1}{(M_\alpha - 1)} \sum_{i=1}^{M_\alpha} \alpha_i^2 , \qquad \sigma_\gamma^2 = \frac{1}{(M_\gamma - 1)} \sum_{j=1}^{M_\gamma} \gamma_j^2 ,$$

$$\sigma_{\tau_\ell}^2 = \frac{1}{(M_{\tau_\ell} - 1)} \sum_{k_\ell=1}^{m} (\tau_{k_\ell}^{(\ell)})^2 \qquad (\ell = 1, 2, \ldots, r) .$$

The assumption of zero means implies that

$$\sum_{i=1}^{M_\alpha} \alpha_i = \sum_{j=1}^{M_\gamma} \gamma_j = \sum_{k_1=1}^{M_{\tau_1}} \tau_{k_1}^{(1)} = \ldots = \sum_{k_r=1}^{M_{\tau_r}} \tau_{k_r}^{(r)} = 0$$

For the mixed model we may assume that the α_i's , γ_j's and $\tau_{k_\ell}^{(\ell)}$'s ($\ell = 1,2,\ldots,r$) are of any of the types described above.

6.11 The Expected Values of the Sums of Squares for Orthogonal Latin Squares

Again we arbitrarily begin with the sums of squares obtained for the Type I model and, since

$$Y_{ijk_1,\ldots,k_r} = \mu + \alpha_i + \gamma_j + \tau_{k_1}^{(1)} + \tau_{k_2}^{(2)} + \cdots + \tau_{k_r}^{(r)} + \varepsilon_{ijk_1,\ldots,k_r} ,$$

$$\bar{Y}_i = \mu + \alpha_i + \bar{\gamma}_. + \bar{\tau}_.^{(1)} + \bar{\tau}_.^{(2)} + \cdots + \bar{\tau}_.^{(r)} + \bar{\varepsilon}_i ,$$

$$\bar{Y}_j = \mu + \bar{\alpha}_. + \gamma_j + \bar{\tau}_.^{(1)} + \bar{\tau}_.^{(2)} + \cdots + \bar{\tau}_.^{(r)} + \bar{\varepsilon}_j ,$$

$$\bar{Y}_{k_1}^{(1)} = \mu + \bar{\alpha}. + \bar{\gamma}. + \tau_{k_1}^{(1)} + \bar{\tau}.^{(2)} + \cdots + \bar{\tau}.^{(\lambda)} + \bar{\varepsilon}_{k_1}^{(1)} ,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\bar{Y}_{k_\lambda}^{(\lambda)} = \mu + \bar{\alpha}. + \bar{\gamma}. + \bar{\tau}.^{(1)} + \bar{\tau}.^{(2)} + \cdots + \tau_{k_\lambda}^{(\lambda)} + \bar{\varepsilon}_{k_\lambda}^{(\lambda)} ,$$

$$\bar{Y} = \mu + \bar{\alpha}. + \bar{\gamma}. + \bar{\tau}.^{(1)} + \bar{\tau}.^{(2)} + \cdots + \bar{\tau}.^{(\lambda)} + \bar{\varepsilon} ,$$

where

$$\bar{\alpha}. = \frac{1}{m}\sum_{i=1}^{m} \alpha_i , \quad \bar{\gamma}. = \frac{1}{m}\sum_{j=1}^{m} \gamma_j , \quad \bar{\tau}.^{(\ell)} = \frac{1}{m}\sum_{k_\ell=1}^{m} \tau_{k_\ell}^{(\ell)} \quad (\ell = 1, 2, \cdots, \lambda),$$

$\bar{\varepsilon}_i.$ is the mean of the m $\varepsilon_{ijk_1,\cdots,k_\lambda}$'s associated with the m observations of the ith row, $\bar{\varepsilon}_j.$ is the mean of the m $\varepsilon_{ijk_1,\cdots,k_\lambda}$'s associated with the m observations of the jth column, $\bar{\varepsilon}_{k_\ell}^{(\ell)}$ is the mean of the m $\varepsilon_{ijk_1,\cdots,k_\lambda}$'s associated with the m values of $Y_{ijk_1,\cdots,k_\lambda}$ which have the given number k as the $(\ell + 2)$nd subscript, and $\bar{\varepsilon}$ is the grand mean of the $\varepsilon_{ijk_1,\cdots,k_\lambda}$'s, we have

$$SSA = m\sum_{i=1}^{m} (\bar{Y}_i. - \bar{Y})^2 = m\sum_{i=1}^{m} (\alpha_i - \bar{\alpha}. + \bar{\varepsilon}_i. - \bar{\varepsilon})^2 ,$$

$$SSC = m\sum_{j=1}^{m} (\bar{Y}_j. - \bar{Y})^2 = m\sum_{j=1}^{m} (\gamma_j - \bar{\gamma}. + \bar{\varepsilon}_j. - \bar{\varepsilon})^2 ,$$

$$SST_1 = m\sum_{k_1=1}^{m} (\bar{Y}_{k_1}^{(1)} - \bar{Y})^2 = m\sum_{k_1=1}^{m} (\tau_{k_1}^{(1)} - \bar{\tau}.^{(1)} + \bar{\varepsilon}_{k_1}^{(1)} - \bar{\varepsilon})^2 ,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$SST_\lambda = m\sum_{k_\lambda=1}^{m} (\bar{Y}_{k_\lambda}^{(\lambda)} - \bar{Y})^2 = m\sum_{k_\lambda=1}^{m} (\tau_{k_\lambda}^{(\lambda)} - \bar{\tau}.^{(\lambda)} + \bar{\varepsilon}_{k_\lambda}^{(\lambda)} - \bar{\varepsilon})^2 ,$$

$$SSE = \sum_{i=1}^{m} \sum_{j=1}^{m} [Y_{ijk_1,\ldots,k_\lambda} - \bar{Y}_{i} - \bar{Y}_{j} - \bar{Y}_{k_1}^{(1)} - \cdots - \bar{Y}_{k_\lambda}^{(\lambda)} + (\lambda+1)\bar{Y}]^2$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} \left[\varepsilon_{ijk_1,\ldots,k_\lambda} - \bar{\varepsilon}_{i} - \bar{\varepsilon}_{j} - \bar{\varepsilon}_{k_1}^{(1)} - \cdots - \bar{\varepsilon}_{k_\lambda}^{(\lambda)} + (\lambda+1)\bar{\varepsilon}\right]^2 .$$

To evaluate the expected values of these sums of squares by Theorem 3.2, we shall need the variances and covariances of the following sets of variables,

$$\alpha_i + \bar{\varepsilon}_i \;, \quad \gamma_j + \bar{\varepsilon}_j \;, \quad \tau_{k_1}^{(1)} + \bar{\varepsilon}_{k_1}^{(1)} \;, \quad \cdots \;, \quad \tau_{k_\lambda}^{(\lambda)} + \bar{\varepsilon}_{k_\lambda}^{(\lambda)} \;.$$

Therefore we must compute

$$E(\alpha_i^2) \,,\; E(\alpha_i \alpha_{i'}) \,,\; E(\gamma_j^2) \,,\; E(\gamma_j \gamma_{j'}) \,,$$

$$E[(\tau_{k_\ell}^{(\ell)})^2] \quad (\ell = 1, 2, \cdots, \lambda) \;,$$

$$E[\tau_{k_\ell}^{(\ell)} \tau_{k_\ell'}^{(\ell)}] \quad (\ell = 1, 2, \cdots, \lambda)$$

in a form which will be valid regardless of the nature of the populations used.

If α_i comes from a Type II population, $E(\alpha_i^2) = \sigma_\alpha^2$, and if it comes from a Type III population,

$$E(\alpha_i^2) = \frac{1}{M_\alpha} \sum_{i=1}^{M_\alpha} \alpha_i^2 = \left(1 - \frac{1}{M_\alpha}\right) \sigma_\alpha^2 \;,$$

which gives the formula for the Type II case if we let $M_\alpha \to \infty$. For the Type II case, $E(\alpha_i \alpha_{i'}) = 0$, and for the Type III case

$$E(\alpha_i \alpha_{i'}) = \sum_{i \neq i'}^{M_\alpha} \frac{\alpha_i \alpha_{i'}}{M_\alpha (M_\alpha - 1)} = - \sum_{i=1}^{M_\alpha} \frac{\alpha_i^2}{M_\alpha (M_\alpha - 1)} = - \frac{\sigma_\alpha^2}{M_\alpha}$$

since

$$0 = \sum_{i,i'}^{M_\alpha} \alpha_i \alpha_{i'} = \sum_{i=1}^{M_\alpha} \alpha_i^2 + \sum_{i \neq i'}^{M_\alpha} \alpha_i \alpha_{i'} \quad .$$

Thus the formula for the Type II case is again included in the Type III case. Similarly for the γ_j's and $\tau_{k_\ell}^{(\ell)}$'s (ℓ = 1, 2,...,r)

$$E(\gamma_j^2) = \left(1 - \frac{1}{M_\gamma}\right) \sigma_\gamma^2 \quad , \quad E(\gamma_j \gamma_{j'}) = -\frac{\sigma_\gamma^2}{M_\gamma} \quad ,$$

$$E[(\tau_{k_\ell}^{(\ell)})^2] = \left(1 - \frac{1}{M_{\tau_\ell}}\right) \sigma_{\tau_\ell}^2 \quad , \quad (\ell = 1, 2, \cdots, r) \quad ,$$

and

$$E[\tau_{k_\ell}^{(\ell)} \tau_{k_\ell'}^{(\ell)}] = -\frac{\sigma_{\tau_\ell}^2}{M_{\tau_\ell}} \quad (\ell = 1, 2, \cdots, r) .$$

To evaluate E(SSA) we let

$$y_i = \alpha_i + \bar{\varepsilon}_i \quad , \quad \bar{y}. = \frac{1}{m} \sum_{i=1}^{m} \alpha_i = \bar{\alpha}. + \bar{\varepsilon} \quad ,$$

so that

$$SSA = m \sum_{i=1}^{m} (y_i - \bar{y}.)^2 \quad ,$$

$$\mu_i = E(y_i) = (1 - \delta_\alpha)\alpha_i \quad , \quad \bar{\mu}. = \frac{1}{m} \sum_{i=1}^{m} \mu_i = 0 \quad ,$$

where $\delta_\alpha = 0$ if the α_i's come from a Type I population and $\delta_\alpha = 1$ otherwise. Also

$$\sigma_i^2 = Var(y_i) = E[\alpha_i + \bar{\varepsilon}_i - E(\alpha_i)]^2$$

$$= E[\alpha_i - E(\alpha_i)]^2 + E(\bar{\varepsilon}_i^2)$$

$$= \delta_\alpha \left(1 - \frac{1}{M_\alpha}\right) \sigma_\alpha^2 + \frac{\sigma^2}{n} \quad ,$$

and

$$\overline{\sigma_{\cdot}}^2 = \frac{1}{m}\sum_{i=1}^{m} \sigma_i^2 = \delta_\alpha \left(1 - \frac{1}{M_\alpha}\right) \sigma_\alpha^2 + \frac{\sigma^2}{m} \quad .$$

Next

$$\lambda = cov(y_i, y_{i'}) = E\{[\alpha_i - E(\alpha_i)][\alpha_{i'} - E(\alpha_{i'})]\}$$
$$= -\delta_\alpha \frac{\sigma_\alpha^2}{M_\alpha} \quad ,$$

$$\overline{\sigma_{\cdot}}^2 - \lambda = \delta_\alpha \sigma_\alpha^2 + \frac{\sigma^2}{m} \quad ,$$

$$E(SSA) = (1-\delta_\alpha) m \sum_{i=1}^{m} \alpha_i^2 + m(m-1)\left(\delta_\alpha \sigma_\alpha^2 + \frac{\sigma^2}{m}\right)$$
$$= (m-1)\sigma^2 + \delta_\alpha m(m-1)\sigma_\alpha^2 + (1-\delta_\alpha) m \sum_{i=1}^{m} \alpha_i^2$$

and

$$E(MSA) = \sigma^2 + \delta_\alpha m \sigma_\alpha^2 + (1-\delta_\alpha)\frac{m}{m-1}\sum_{i=1}^{m} \alpha_i^2 \quad .$$

Similarly

$$E(MSC) = \sigma^2 + \delta_\gamma m \sigma_\gamma^2 + (1-\delta_\gamma)\frac{m}{m-1}\sum_{j=1}^{m} \gamma_j^2 \quad ,$$

$$E(MST_1) = \sigma^2 + \delta_{\tau_1} m \sigma_{\tau_1}^2 + (1-\delta_{\tau_1})\frac{m}{m-1}\sum_{k_1=1}^{m} (\tau_{k_1}^{(1)})^2 \quad ,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$E(MST_\lambda) = \sigma^2 + \delta_{\tau_\lambda} m \sigma_{\tau_\lambda}^2 + (1-\delta_{\tau_\lambda})\frac{m}{m-1}\sum_{k_\lambda=1}^{m} (\tau_{k_\lambda}^{(\lambda)})^2 \quad .$$

Finally, by the theory of Chapter II, we know that SSE is distributed as $\chi^2\sigma^2$ with (m-1)(m-r-1) d.f. . Hence

$$E\left(\frac{SSE}{\sigma^2}\right) = (m-1)(m-r-1)$$

and

$$E(MSE) = \sigma^2 .$$

If we define

$$\sigma_\alpha^2 = \frac{1}{m-1}\sum_{i=1}^{m}\alpha_i^2 , \qquad \sigma_\gamma^2 = \frac{1}{m-1}\sum_{j=1}^{m}\gamma_j^2 ,$$

$$\sigma_{\tau_\ell}^2 = \frac{1}{m-1}\sum_{k_\ell=1}^{m}(\tau_{k_\ell}^{(\ell)})^2 \qquad (\ell = 1, 2, \cdots, r) ,$$

for the Type I populations, the above results are summarized in table 6.6.

The following formulas are more convenient for computation:

$$SSA = m\sum_{i=1}^{m}(\bar{Y}_{i\cdot} - \bar{Y})^2 = \sum_{i=1}^{m}\frac{Y_{i\cdot}^2}{m} - \frac{Y^2}{m^2} ,$$

$$SSC = m\sum_{j=1}^{m}(\bar{Y}_{\cdot j} - \bar{Y})^2 = \sum_{j=1}^{m}\frac{Y_{\cdot j}^2}{m} - \frac{Y^2}{m^2} ,$$

$$SST_1 = m\sum_{k_1=1}^{m}(\bar{Y}_{k_1}^{(1)} - \bar{Y})^2 = \sum_{k_1=1}^{m}\frac{(Y_{k_1}^{(1)})^2}{m} - \frac{Y^2}{m} ,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$SST_r = m\sum_{k_r=1}^{m}(\bar{Y}_{k_r}^{(r)} - \bar{Y})^2 = \sum_{k_r=1}^{m}\frac{(Y_{k_r}^{(r)})^2}{m} - \frac{Y^2}{m} ,$$

$$\sum_{i=1}^{m}\sum_{j=1}^{m}(Y_{ijk_1,\ldots,k_r} - \bar{Y})^2 = \sum_{i,j}(Y_{ijk_1,\cdots,k_r}^2) - \frac{Y^2}{m^2} ,$$

where $Y = \sum_{i,j} Y_{ijk_1,\cdots,k_r}$.

We find SSE by subtraction.

TABLE 6.6

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	E(MS)
Rows	m-1	$SSA = m\sum_{i=1}^{m}(\bar{Y}_{i\cdot} - \bar{Y})^2$	$MSA = \frac{SSA}{m-1}$	$\sigma^2 + m\sigma_{\alpha}^2$
Columns	m-1	$SSC = m\sum_{j=1}^{m}(\bar{Y}_{\cdot j} - \bar{Y})^2$	$MSC = \frac{SSC}{m-1}$	$\sigma^2 + m\sigma_{\gamma}^2$
Treatments	m-1	$SST_1 = m\sum_{k_1=1}^{m}(\bar{Y}_{k_1}^{(1)} - \bar{Y})^2$	$MST_1 = \frac{SST_1}{m-1}$	$\sigma^2 + m\sigma_{\tau_1}^2$
	...	....	....	....
	m-1	$SST_r = m\sum_{k_r=1}^{m}(\bar{Y}_{k_r}^{(r)} - \bar{Y})^2$	$MST_r = \frac{SST_r}{m-1}$	$\sigma^2 + m\sigma_{\tau_r}^2$
Error	(m-1)(m-r-1)	$SSE = \sum_{i=1}^{m}\sum_{j=1}^{m}\left[Y_{ijk_1,\ldots,k_r} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} - \bar{Y}_{k_1}^{(1)} - \cdots - \bar{Y}_{k_r}^{(r)} + (r+1)\bar{Y}\right]^2$	$MSE = \frac{SSE}{(m-1)(m-r-1)}$	σ^2
Total	m^2-1	$\sum_{i=1}^{m}\sum_{j=1}^{m}(Y_{ijk_1,\ldots,k_r} - \bar{Y})^2$		

6.12 Distributions of the Sums of Squares for Orthogonal Latin Squares

Corresponding to the hypotheses

$$H_1: \alpha_i = 0 \; , \; H_2: \gamma_j = 0 \; , \; H_3: \tau_{k_1}^{(1)} = 0 \; , \ldots , \; H_{r+2}: \tau_{k_r}^{(r)} = 0 \; ,$$

we have the hypotheses

$$\sigma_\alpha = 0 , \quad \sigma_\gamma = 0 , \quad \sigma_{\tau_1} = 0 , \quad \cdots , \quad \sigma_{\tau_r} = 0 ,$$

if the corresponding variables are from other than a Type I population. Then, since the populations have zero means, it follows that the corresponding variables are equal to zero.

In the case of the Type I model, the theory of Chapter II implies that

$$\bar{y}_{i\cdot} - \bar{y} , \quad \bar{y}_{\cdot j} - \bar{y} , \quad \bar{y}_{k_1}^{(1)} - \bar{y} , \quad \cdots , \quad \bar{y}_{k_r}^{(r)} - \bar{y} ,$$

$$(i, j, k_1, \ldots, k_r = 1, 2, \ldots, m-1)$$

are distributed independently of SSE. Therefore any function of these expressions is distributed independently of SSE and, in particular, this holds for

$$\bar{Y}_{m\cdot} - \bar{Y} \quad (\underline{m}\text{th row}) ,$$

$$\bar{Y}_{\cdot m} - \bar{Y} \quad (\underline{m}\text{th column}) ,$$

$$\bar{Y}_m^{(1)} - \bar{Y} , \; \ldots , \; \bar{Y}_m^{(r)} - \bar{Y} \; .$$

These results were obtained for the model

$$y_{ijk_1,\ldots,k_r} = \mu + \alpha_i + \gamma_j + \tau_{k_1}^{(1)} + \tau_{k_2}^{(2)} + \cdots + \tau_{k_r}^{(r)} + \varepsilon_{ijk_1,\ldots,k_r}$$

and they hold for the particular case where

$$y_{ijk_1,\ldots,k_r} = \varepsilon_{ijk_1,\ldots,k_r} \; .$$

Hence

$$\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon} , \quad \bar{\varepsilon}_{\cdot j} - \bar{\varepsilon} , \quad \bar{\varepsilon}_{k_1}^{(1)} - \bar{\varepsilon} , \quad \cdots , \quad \bar{\varepsilon}_{k_r}^{(r)} - \bar{\varepsilon} ,$$

are distributed independently of SSE.

We shall now show that these expressions in the $\varepsilon_{ijk_1,\cdots,k_r}$'s are independent of each other. Since they have normal distributions, it is sufficient to prove that

$$\operatorname{cov}(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon},\ \bar{\varepsilon}_{j\cdot} - \bar{\varepsilon}) = \operatorname{cov}(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon},\ \bar{\varepsilon}^{(l)}_{k_l} - \bar{\varepsilon})$$

$$= \operatorname{cov}(\bar{\varepsilon}_{j\cdot} - \bar{\varepsilon},\ \bar{\varepsilon}^{(l)}_{k_l} - \bar{\varepsilon}) = 0 \qquad (l = 1, 2, \cdots, r) .$$

Since

$$E(\bar{\varepsilon}_{i\cdot}\ \bar{\varepsilon}_{j\cdot}) = \sigma^2/m^2 ,$$

$$E(\bar{\varepsilon}_{i\cdot}\ \bar{\varepsilon}) = \frac{m}{m^3}\sigma^2 = \sigma^2/m^2 ,$$

$$E(\bar{\varepsilon}_{j\cdot}\ \bar{\varepsilon}) = \frac{m}{m^3}\sigma^2 = \sigma^2/m^2 ,$$

$$E(\bar{\varepsilon}^2) = \sigma^2/m^2 ,$$

we have

$$\operatorname{cov}(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon},\ \bar{\varepsilon}_{j\cdot} - \bar{\varepsilon}) = \sigma^2\left(\frac{1}{m^2} - \frac{1}{m^2} - \frac{1}{m^2} + \frac{1}{m^2}\right) = 0 .$$

Also

$$E(\bar{\varepsilon}_{i\cdot}\ \bar{\varepsilon}^{(l)}_{k_l}) = \sigma^2/m^2 \qquad (l = 1, 2, \cdots, r) ,$$

$$E(\bar{\varepsilon}\ \bar{\varepsilon}^{(l)}_{k_l}) = \frac{m}{m^3}\sigma^2 = \sigma^2/m^2 \qquad (l = 1, 2, \cdots, r) ,$$

so that

$$\operatorname{cov}(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon},\ \bar{\varepsilon}^{(l)}_{k_l} - \bar{\varepsilon}) = 0 \qquad (l = 1, 2, \cdots, r) .$$

Similarly

$$\operatorname{cov}(\bar{\varepsilon}_{j\cdot} - \bar{\varepsilon},\ \bar{\varepsilon}^{(l)}_{k_l} - \bar{\varepsilon}) = 0 \qquad (l = 1, 2, \cdots, r) .$$

In section 6.9 we tested the hypotheses H_1, H_2,..., H_{r+2}, for the Type I model by the statistics

$$F_1 = \frac{MSA}{MSE}, \quad F_2 = \frac{MSC}{MSE}, \quad F_3 = \frac{MST_1}{MSE}, \ldots, \quad F_{r+2} = \frac{MST_r}{MSE},$$

respectively, since subject to the corresponding hypothesis,

$$SSA = m\sum_{i=1}^{m}(\bar{\varepsilon}_i - \bar{\varepsilon})^2, \qquad SSC = m\sum_{j=1}^{m}(\bar{\varepsilon}_j - \bar{\varepsilon})^2,$$

$$SST_1 = m\sum_{k_1=1}^{m}(\bar{\varepsilon}_{k_1}^{(1)} - \bar{\varepsilon})^2, \quad \ldots \quad, SST_r = m\sum_{k_r=1}^{m}(\bar{\varepsilon}_{k_r}^{(r)} - \bar{\varepsilon})^2,$$

$$SSE = \sum_{i=1}^{m}\sum_{j=1}^{m}\left[\varepsilon_{ijk_1,\ldots,k_r} - \bar{\varepsilon}_i - \bar{\varepsilon}_j - \bar{\varepsilon}_{k_1}^{(1)} - \cdots - \bar{\varepsilon}_{k_r}^{(r)} + (r+1)\bar{\varepsilon}\right]^2,$$

are independently distributed as $\chi^2\sigma^2$, the first r+2 sums with m-1 d.f. and the last sum with (m-1)(m-r-1) d.f. .

We shall now determine what tests can be made when we are not dealing with a Type I model. In section 6.11 we saw that

$$SSA = m\sum_{i=1}^{m}(\alpha_i - \bar{\alpha}. + \bar{\varepsilon}_i - \bar{\varepsilon})^2,$$

$$SSC = m\sum_{j=1}^{m}(\gamma_j - \bar{\gamma}. + \bar{\varepsilon}_j - \bar{\varepsilon})^2,$$

$$SST_1 = m\sum_{k_1=1}^{m}(\tau_{k_1}^{(1)} - \bar{\tau}.^{(1)} + \bar{\varepsilon}_{k_1}^{(1)} - \bar{\varepsilon})^2,$$

$$\cdots\cdots\cdots$$

$$SST_r = m\sum_{k_r=1}^{m}(\tau_{k_r}^{(r)} - \bar{\tau}.^{(r)} + \bar{\varepsilon}_{k_r}^{(r)} - \bar{\varepsilon})^2.$$

Subject to H_1, H_2,..., H_{r+2}, the sums of squares SSA, SSC, SST_1, ...,SST_r reduce to the corresponding expressions for the Type I model and the same tests apply.

BIBLIOGRAPHY

Attridge, R.F., Linear Regression and Multiple Classification Designs. Hamilton: unpublished thesis, 1952, p. 63.

Bedrosian, Peter, Orthogonal Latin Squares and Incomplete Balanced Block Designs. Hamilton: unpublished thesis, 1953, p. 90.

Bennett and Franklin, Statistical Analysis in Chemistry and the Chemical Industry. New York: John Wiley & Sons, 1954, p. 320.

Mann, H.B., Analysis and Design of Experiments. New York: Dover Publications, 1949, p. 39.

Mood, A.M., Introduction to the Theory of Statistics. New York: McGraw-Hill Co., 1950, p. 159.

Patnaik, P.B., The Non-Central χ^2- and F-Distributions and their Applications, Biometrika, 36 (1949), p. 202.

Satterthwaite, F.E., An Approximate Distribution of Estimates of Variance Components. Biometrics Bull, 2 (1946), pp. 110-114.

Schreier and Sperner, Modern Algebra and Matrix Theory. New York: Chelsea Publishing Company, 1951, p. 141.

Tukey, J.W., Dyadic Analysis of Variance, Human Biology, 21 (1949), pp. 65-110.

Welch, B.L., The Specification of Rules for Rejecting too Variable a Product. J. Royal Stat. Soc. (Supp.), 3 (1936), p. 29.