SPLIT PLOT DESIGNS

by

MARCIAL PEREZ ESTOLANO

B.S.A., University of the Philippines at Los Baños, 1976

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1985

Approved by:

James J. Higgins
Major Professor

TABLE OF CONTENTS

INTRODUCTION

The split-plot design is used in two-factor experiments where one factor requires larger experimental units than the other. Suppose that we wish to compare several levels of furnaces and moulds for making alloys. A fairly large amount of material is made from each furnace, and the material is then poured (split) into different types of mould. With, say, 3 levels of furnaces (A) and 4 levels of moulds (B), a split-plot design with 3 replications might look as follows, after randomization. Each replication is made up of 3 large units, each unit representing a type of furnace, and each large unit is split into 4 'plots', each plot representing a type of mould.

	Replication 1		Replication 2		Replication 3
a_2	b_1 b_3 b_4 b_2	a_3	b_3 b_4 b_2 b_1	a_1	b_2 b_4 b_1 b_3
a_3	b_2 b_4 b_1 b_3	a_1	b_1 b_3 b_4 b_2	a_2	b_4 b_1 b_3 b_2
a_1	b_4 b_1 b_3 b_2	a_2	b_2 b_1 b_3 b_4	a_3	b_3 b_2 b_4 b_1

In field experiments, this design can arise when an extra factor is introduced into the experiment by dividing each block of a field into a number of parts. Suppose that the experiment is planned originally to test factor A with 3 levels. The division of each unit for A into 4 parts permits the inclusion of an extra factor B at four levels. Within each unit for A the four levels of B are allocated at random to the four plots. The plan after randomization might appear as the above plan.

In the terminology of split designs, the larger units are called main plots and the smaller units are called sub-plots. The randomization takes place in two stages --- the allocation of the whole plot treatments to the whole plots, and the allocation of the sub-plot treatments to the sub-plots within each main plot (random assignment). In the classical split plot design the whole plots are arranged in Randomized Complete Blocks. Other popular designs for the whole plots are Completely Random and Latin Squares. The development below will be for the Randomized Complete Block arrangement of whole plots.

The mathematical model for this experiment is

$$y_{ijk} = \mu + \rho_i + \alpha_j + \delta_{ij} + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \qquad (1)$$

$$i = 1,2,\ldots,r \quad j = 1,2,\ldots,a \quad k = 1,2,\ldots,b$$

where μ = overall mean

ρ_i = block/replication effect

α_j = whole plot treatment effect

δ_{ij} = residual main plot effect [error (a)]

β_k = sub-plot treatment effect

$(\alpha\beta)_{jk}$ = whole plot treatment*sub-plot treatment interaction effect

ε_{ijk} = random sub-plot error [error (b)]

In the analysis of this type of design, sources which are part of the whole-plot variation are usually grouped separately from those which are part of the sub-plot variation.

The model of equation (1) can then be expressed in terms of each size of experimental unit:

$$\begin{aligned} y_{ijk} &= \mu + \rho_i + \alpha_j + \delta_{ij} && \text{] whole plot of the model} \\ &\quad + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} && \text{] sub-plot of the model} \end{aligned}$$

II. DESIGN CONSIDERATION

(1) Classical Split Plot Design.

Even when one factor does not require larger plots than the other, the split-plot design still may be appropriate. We might have started the experiment to compare the main plot treatments only. The sub-plot treatments might have arisen as an afterthought, or might even be suggested by the experiment already in progress. Another situation is where we are not too interested in one factor (we might already have sufficient information on the effects of this factor) and we are much more interested in the effect of the other factor as well as interactions. In the simple Randomized Complete Block design, any two treatments are compared equally precisely while in the split-plot design, we sacrifice precision of the comparison among the main plot treatments to increase the precision of the comparison among the sub-plot treatments. This is intuitively clear, since the sub-plot treatments are compared on more homogeneous experimental units than the main plot treatments. Also, the sub-plot treatments are more highly replicated than the main plot treatments.

Several examples of the use of this design are as follows:

(i) Comparison of several recipes of cake mix and baking temperatures. A sufficiently large amount of cake mix is made using each recipe. The batter from each recipe is then poured into several cake pans for baking at different temperatures.

(ii) Comparison of several milking machines (main plot) and several methods of pasteurizing the milk (sub-plot).

(iii) In Phytopathology, plants make natural plot and leaves within a plant form the sub-plots.

(iv) In organoleptic testing, we may be comparing several brands of orange juice. A can from each brand may be split into several aliquots for judging by a taste panel. The replication of main plots treatments will be provided by conducting the tasting on each of several days.

In consideration of the relative merits of randomized complete block design and split plot design, Cochran and Cox (8) pointed out that the following points are relevant.

(i) With the use of split-plot design, usually the B and the AB effects are estimated more precisely than the A effects. Moreover, the number of degrees of freedom available for the experimental error mean square is smaller for the whole plot comparisons than for the sub-plot comparisons.

(ii) The increase precision on B and AB effects is obtained by the sacrifice of precision on A effects using the split-plot design. However, for tests of significance or the construction of confidence limits, the randomized complete block design holds a slight advantage on the average since it provides more degrees of freedom for the estimate of the single error variance (assuming that experimental error can be controlled equally well in both designs).

(iii) The chief practical advantage of the split-plot arrangement is that it enables factors that require relatively large amounts of materials and factors that require only small amounts to be combined in the same experiment. If the experiment is planned to investigate the first type of factor, so that large amounts of material are going to be used anyway, factors of the second type can often be included at very little extra costs and some additional information can be obtained very cheaply.

(2) Assumptions

Assume that we have r number of blocks, a levels of factor A, b levels of factor B and one observation per cell. The model for the yield of the ith block from the jth level of factor A on the kth level of factor B is given by model (1). To carry out the analysis of variance, we make several assumptions on the parameters of the given model. The assumptions and restrictions imposed on the model are:

1) $\delta_{ij} \sim N(0,\sigma_d^2)$

2) $\varepsilon_{ijk} \sim N(0,\sigma_e^2)$

3) δ_{ij} and ε_{ijk} are independent

4) Factor A and Factor B are fixed

5) Blocks are random

6) $\sum_j \alpha_j = \sum_k \beta_k = \sum_j (\alpha\beta)_{jk} = \sum_k (\alpha\beta)_{jk} = 0$

The total sum of squares of the deviations of y_{ijk} from the overall mean can be subdivided into six independent parts by means of the following identity.

$$\sum_{ijk}\sum\sum (y_{ijk} - \bar{y}_{...})^2$$

$$= \sum_{ijk}\sum\sum [(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})$$

$$+ (\bar{y}_{..k} - \bar{y}_{...}) + (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})$$

$$+ (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{...})]^2$$

$$= \sum_{ijk}\sum\sum (\bar{y}_{i..} - \bar{y}_{...})^2 + \sum_{ijk}\sum\sum (\bar{y}_{.j.} - \bar{y}_{...})^2 + \sum_{ijk}\sum\sum (\bar{y}_{..k} - \bar{y}_{...})^2$$

$$+ \sum_{ijk}\sum\sum (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{ijk}\sum\sum (y_{.jk} - y_{.j.} - y_{..k} + y_{...})^2$$

$$+ \sum_{ijk}\sum\sum (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{.j.})^2 + \text{zero cross-products.}$$

$$= ab\sum_{i} (\bar{y}_{i..} - \bar{y}_{...})^2 + br\sum_{j} (\bar{y}_{.j.} - \bar{y}_{...})^2 + ar\sum_{k} (\bar{y}_{..k} - \bar{y}_{...})^2$$

$$+ b\sum_{ij}\sum (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + r\sum_{jk}\sum (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2$$

$$+ \sum_{ijk}\sum\sum (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{.j.})^2$$

The quantity on the left hand side of the equation is nothing but the total sum of squares (SST) and the quantities on the right hand side are the subdivision or partitioning of the total sum of squares into its components, namely the sum of squares due to blocks (SSR), sum of squares due to wholeplot treatment (SSA), sum of squares due to whole plot error (SSE(a)), sum of squares due to sub-plot treatment (SSB), sum of squares due to the interaction of whole plot treatment and sub-plot treatment (SSAB) and the sum of squares due to random error (SSE(b)). Hence, we have the relation

$$SST = SSR + SSA + SSE(a) + SSB + SSAB + SSE(b). \qquad (2)$$

Likewise, the total degrees of freedom which is (abr-1) is also partitioned into its components, that is,

$$(abr-1) = (r-1) + (a-1) + (r-1)(a-1) + (b-1) + (a-1)(b-1) + a(b-1)(r-1). \quad (3)$$

Corresponding to each source of variation, the mean square is obtained by dividing the sum of squares by its corresponding degrees of freedom.

An analysis of variance table for split plot design is shown in Table 1, where A denotes the whole plot treatment with a levels, B denotes the sub-plot treatment with b levels, and r denotes the number of replications.

Table 1. Analysis of Variance Table for Split-Plot Design

Source	df	SS	MS	E(MS)
Replication	r-1	SSR	MSR	$\sigma_e^2 + b\sigma_d^2 + ab\sigma_r^2$
A	a-1	SSA	MSA	$\sigma_e^2 + b\sigma_d^2 + rb\Sigma\,\alpha_j^2/(a-1)$
Error(a)	(a-1)(r-1)	SSE(a)	MSE(a)	$\sigma_e^2 + b\sigma_d^2$
B	b-1	SSB	MSB	$\sigma_e^2 + ar\Sigma\,\beta_k^2/(b-1)$
A*B	(a-1)(b-1)	SSAB	MSAB	$\sigma_e^2 + r\Sigma\Sigma\,(\alpha\beta)_{jk}^2/(a-1)(b-1)$
Error(b)	a(b-1)(r-1)	SSE(b)	MSE(b)	σ_e^2
Total	abr-1	SST		

(3) Derivation of Expected Mean Squares.

The expected mean squares are derived algebraically by using the model assumptions, that is, we use the mean squares obtained from the analysis of the given model and evaluate the average value for each source of variation. They are as follows:

a) Expected mean squares of Factor A.

$E(MSA) = E\ [rb\sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2/(a-1)]$, where the quantities inside the parethesis sign can be expressed as:

$$\bar{y}_{.j.} = \mu + \bar{\rho}_{.} + \alpha_j + \bar{\delta}_{.j} + \bar{\varepsilon}_{.j.}$$

$$\bar{y}_{...} = \mu + \bar{\rho}_{.} + \bar{\delta}_{..} + \bar{\varepsilon}_{...}$$

Substituting the terms we get

$$(\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$= (\alpha_j + \bar{\delta}_{.j} + \bar{\varepsilon}_{.j.} - \bar{\delta}_{..} - \bar{\varepsilon}_{...})^2$$

$$= [\alpha_j + (a\sum_i \delta_{ij} - \sum\sum_{ij} \delta_{ij})/ar + (a\sum\sum_{ik} \varepsilon_{ijk} - \sum\sum\sum_{ijk} \varepsilon_{ijk})/abr]^2$$

$$= \alpha_j^2 + (a^2\sum_i \delta_{ij}^2 - 2a\sum_i \delta_{ij}^2 + \sum\sum_{ij} \delta_{ij}^2)/a^2r^2$$

$$+ (a^2\sum\sum_{ik} \varepsilon_{ijk}^2 - 2a\sum\sum_{ik} \varepsilon_{ijk}^2 + \sum\sum\sum_{ijk} \varepsilon_{ijk}^2)/a^2b^2r^2$$

+ cross products of independent mean zero terms.

Taking the expectations, we get

$$E(\bar{y}_{.j.} - \bar{y}_{...})^2 = \alpha_j^2 + \sigma_d^2(a^2r - 2ar + ar)/a^2r^2 + \sigma_e^2(a^2rb - 2arb + arb)/a^2b^2r^2$$

$$= \alpha_j^2 + \sigma_d^2(a-1)/ar + \sigma_e^2(a-1)/abr$$

Thus, $E[rb\sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2/(a-1)] = \sigma_e^2 + b\sigma_d^2 + rb\sum \alpha_j^2/(a-1)$

b) Expected mean squares for main plot residual term

$E[MSE(a)] = E[b\sum\sum_{ij} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2/(a-1)(r-1)]$, where each

quantites inside the parethesis sign can be expressed as

$$\bar{y}_{ij.} = \mu + \rho_i + \alpha_j + \delta_{ij} + \bar{\varepsilon}_{ij.}$$

$$\bar{y}_{i..} = \mu + \rho_i + \bar{\delta}_{i.} + \bar{\varepsilon}_{i..}$$

$$\bar{y}_{.j.} = \mu + \bar{\rho}_{.} + \alpha_j + \bar{\delta}_{.j} + \bar{\varepsilon}_{.j.}$$

$$\bar{y}_{...} = \mu + \bar{\rho}_{.} + \bar{\delta}_{..} + \bar{\varepsilon}_{...}$$

Substituting the terms we get

$$(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$= [(\delta_{ij} - \bar{\delta}_{i.} - \bar{\delta}_{.j} + \bar{\delta}_{..}) + (\bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{...})]^2$$

$$= [(ar\delta_{ij} - r\sum_j \delta_{ij} - a\sum_i \delta_{ij} + \sum\sum_{ij} \delta_{ij})/ar$$

$$+ (ar\sum_k \varepsilon_{ijk} - r\sum\sum_{jk} \varepsilon_{ijk} - a\sum\sum_{ik} \varepsilon_{ijk} + \sum\sum\sum_{ijk} \varepsilon_{ijk})/abr]^2$$

$$= [(a^2r\delta_{ij}^2 + r^2\sum_j \delta_{ij}^2 + a^2\sum_i \delta_{ij}^2 + \sum\sum_{ij} \delta_{ij}^2 - 2ar^2\delta_{ij}^2 - 2a^2r\delta_{ij}^2 + 2ar\delta_{ij}^2$$

$$+ 2ar\delta_{ij}^2 - 2r\sum_j \delta_{ij}^2 - 2a\sum_i \delta_{ij}^2)/a^2r^2] + [(a^2r^2\sum_k \varepsilon_{ijk}^2 + r^2\sum\sum_{jk} \varepsilon_{ijk}^2$$

$$+ a^2\sum\sum_{ik} \varepsilon_{ijk}^2 + \sum\sum\sum_{ijk} \varepsilon_{ijk}^2 - 2a^2r\sum_k \varepsilon_{ijk}^2 - 2ar^2\sum_k \varepsilon_{ijk}^2 + 2ar\sum_k \varepsilon_{ijk}^2$$

$$+ 2ar\sum_k \varepsilon_{ijk}^2 - 2r\sum\sum_{jk} \varepsilon_{ijk}^2 - 2a\sum\sum_{ik} \varepsilon_{ijk}^2)/a^2b^2r^2]$$

+ cross products of independent mean zero terms.

Taking the expectations

$$E(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$= \sigma_d^2(a^2r^2 + r^2a + ra^2 + ra - 2ar^2 - 2a^2r + 2ar + 2ar - 2ar - 2ar)/a^2r^2$$

$$+ \sigma_e^2(a^2r^2b + r^2ab + a^2rb + abr - 2ar^2b - 2a^2rb + 2arb + 2arb$$

$$- 2arb - 2arb)/a^2b^2r^2$$

$$= \sigma_d^2(a-1)(r-1)/ar + \sigma_e^2(a-1)(r-1)/abr$$

Thus, $E[b\sum_{ij}\sum (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2/(a-1)(r-1)] = \sigma_e^2 + b\sigma_d^2$

c) Expected mean squares for Factor B.

$E(MSB) = E[ar\sum_k (\bar{y}_{..k} - \bar{y}_{...})^2/b-1]$, were each quantities inside the parenthesis sign can be expressed as follows

$$\bar{y}_{..k} = \mu + \bar{\rho}_{.} + \bar{\delta}_{..} + \beta_k + \bar{\varepsilon}_{..k}$$

$$\bar{y}_{...} = \mu + \bar{\rho}_{..} + \bar{\delta}_{..} + \bar{\varepsilon}_{...}$$

Substituting this quantities, we get

$$(\bar{y}_{..k} - \bar{y}_{...})^2 = (\beta_k + \bar{\varepsilon}_{..k} - \bar{\varepsilon}_{...})^2$$

$$= [\beta_k + (b\sum_{ij}\sum \varepsilon_{ijk} - \sum\sum_{ijk}\sum \varepsilon_{ijk})/abr]^2$$

$$= \beta_k^2 + (b^2\sum_{ij}\sum \varepsilon_{ijk}^2 + \sum\sum_{ijk}\sum \varepsilon_{ijk}^2 - 2b\sum_{ij}\sum \varepsilon_{ijk}^2)/a^2b^2r^2$$

Taking the expectations

$$E(\bar{y}_{..k} - \bar{y}_{...})^2 = \beta_k^2 + \sigma_e^2(ab^2r + abr - 2abr)/a^2b^2r^2$$

$$= \beta_k^2 + \sigma_e^2(b-1)/abr$$

Thus, $E[ar\sum_k (\bar{y}_{..k} - \bar{y}_{...})^2/b-1] = \sigma_e^2 + ar\sum \beta_k^2/(b-1)$

d) Expected mean squares of the interaction

$E(MSAB) = E[r\sum\sum_{jk} (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2/(a-1)(b-1)]$, where each

quantites inside the parethesis sign can be expressed as follows

$$\bar{y}_{.jk} = \mu + \bar{\rho}_{.} + \alpha_j + \bar{\delta}_{.j} + \beta_k + (\alpha\beta)_{jk} + \bar{\varepsilon}_{.jk}$$

$$\bar{y}_{.j.} = \mu + \bar{\rho}_{.} + \alpha_j + \bar{\delta}_{.j} + \bar{\varepsilon}_{.j.}$$

$$\bar{y}_{..k} = \mu + \bar{\rho}_{.} + \bar{\delta}_{..} + \beta_k + \bar{\varepsilon}_{..k}$$

$$\bar{y}_{...} = \mu + \bar{\rho}_{.} + \bar{\delta}_{..} + \bar{\varepsilon}_{...}$$

Substituting the terms, we get

$$(\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2$$

$$= [(\alpha\beta)_{jk} + (\bar{\varepsilon}_{.jk} - \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{..k} + \bar{\varepsilon}_{...})]^2$$

$$= [(\alpha\beta)_{jk} + (ab\sum_i \varepsilon_{ijk} - a\sum\sum_{ik} \varepsilon_{ijk} - b\sum\sum_{ij} \varepsilon_{ijk} + \sum\sum\sum_{ijk} \varepsilon_{ijk})/abr]^2$$

$$= (\alpha\beta)_{jk}^2 + [a^2b^2\sum_i \varepsilon_{ijk}^2 + a^2\sum\sum_{ik} \varepsilon_{ijk}^2 + b^2\sum\sum_{ij} \varepsilon_{ijk}^2 + \sum\sum\sum_{ijk} \varepsilon_{ijk}^2$$

$$- 2a^2b\sum_i \varepsilon_{ijk}^2 - 2ab^2\sum_i \varepsilon_{ijk}^2 + 2ab\sum_i \varepsilon_{ijk}^2 + 2ab\sum_i \varepsilon_{ijk}^2$$

$$- 2a\sum\sum_{ik} \varepsilon_{ijk}^2 - 2b\sum\sum_{ij} \varepsilon_{ijk}^2]/a^2b^2r^2$$

+ cross products of independent mean zero terms.

Taking expectations

$$E(\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2$$

$$= (\alpha\beta)^2_{jk} + \sigma^2_e(a^2b^2r + a^2br + ab^2r + abr - 2a^2br - 2br^2 + 2abr$$

$$+ 2abr - 2abr - 2abr)/a^2b^2r^2$$

$$= (\alpha\beta)^2_{jk} + \sigma^2_e(a-1)(b-1)/abr$$

Thus, $E[r\sum_{jk}\sum (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2/(a-1)(b-1)]$

$$= \sigma^2_e + r\sum\sum(\alpha\beta)^2_{jk}/(a-1)(b-1)$$

e) Expected mean squares of random error

$E[MSE(b)] = E[\sum_{ijk}\sum\sum (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{.j.})^2/a(b-1)(r-1)]$, where each

quantities inside the parethesis sign can be expressed as follows

$$y_{ijk} = \mu + \rho_i + \alpha_j + \delta_{ij} + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}$$

$$\bar{y}_{ij.} = \mu + \rho_i + \alpha_j + \delta_{ij} + \bar{\varepsilon}_{ij.}$$

$$\bar{y}_{.jk} = \mu + \bar{\rho}_. + \bar{\delta}_{.j} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \bar{\varepsilon}_{.jk}$$

$$\bar{y}_{.j.} = \mu + \bar{\rho}_. + \bar{\delta}_{.j} + \alpha_j + \bar{\varepsilon}_{.j.}$$

Substituting the terms, we get

$$(y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{.j.})^2$$

$$= (\varepsilon_{ijk} - \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{.jk} + \bar{\varepsilon}_{.j.})^2$$

$$= [(br\varepsilon_{ijk} - r\sum_k \varepsilon_{ijk} - b\sum_i \varepsilon_{ijk} + \sum_{ik}\sum \varepsilon_{ijk})/br]^2$$

$$= (b^2r^2\varepsilon^2_{ijk} + r^2\sum_k \varepsilon^2_{ijk} + b^2\sum_i \varepsilon^2_{ijk} + \sum_{ik}\sum \varepsilon^2_{ijk} - 2br^2\varepsilon^2_{ijk} - 2b^2r\varepsilon^2_{ijk}$$

$$+ 2br\varepsilon^2_{ijk} + 2br\varepsilon^2_{ijk} - 2r\sum_k \varepsilon^2_{ijk} - 2b\sum_i \varepsilon^2_{ijk})/b^2r^2$$

+ cross products of independent mean zero terms.

Taking the expectations

$$E(y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{.j.})^2 = \sigma^2_e(b^2r^2 + br^2 + b^2r + b_2r_2 - 2b^2r - 2br^2 + 2br$$

$$+ 2br - 2br - 2br)/br$$

$$= \sigma^2_e(b-1)(r-1)/br$$

Thus, $E[\sum_{ijk}\sum\sum (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{.j.})^2/a(b-1)(r-1)] = \sigma^2_e$

f) Expected mean squares of the blocks

$E(MSR) = E[ab\sum_i (\bar{y}_{i..} - \bar{y}_{...})^2/(r-1)]$, where each quantites inside the parethesis sign can be expressed as follows

$$\bar{y}_{i..} = \mu + \rho_i + \bar{\delta}_{i.} + \bar{\varepsilon}_{i..}$$

$$\bar{y}_{...} = \mu + \bar{\rho}_{.} + \bar{\delta}_{..} + \bar{\varepsilon}_{...}$$

Substituting the terms, we get

$$(\bar{y}_{i..} - \bar{y}_{...})^2$$

$$= (\rho_i - \bar{\rho}_{.} + \bar{\delta}_{i.} - \bar{\delta}_{..} + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2$$

$$= [(r\rho_i - \sum_i \rho_i)/r + (r\sum_j \delta_{ij} - \sum\sum_{ij} \delta_{ij})/ar + (r\sum\sum_{jk} \varepsilon_{ijk} - \sum\sum\sum_{ijk} \varepsilon_{ijk})/abr]^2$$

$$= (r^2\rho_i^2 + \sum_i \rho_i^2 - 2r\rho_i^2)/r^2 + (r^2\sum_j \delta_{ij}^2 + \sum\sum_{ij} \delta_{ij}^2 - 2r\sum_j \delta_{ij}^2)/a^2r^2$$

$$+ (r^2\sum\sum_{jk} \varepsilon_{ijk}^2 + \sum\sum\sum_{ijk} \varepsilon_{ijk}^2 - 2r\sum\sum_{jk} \varepsilon_{ijk}^2)/a^2b^2r^2$$

+ cross products of independent mean zero terms.

Taking the expectations

$$E(\bar{y}_{i..} - \bar{y}_{...})^2$$

$$= \sigma_r^2(r^2 + r - 2r)/r^2 + \sigma_d^2(ar^2 + ar - 2ar)/a^2r^2$$

$$+ \sigma_e^2(abr^2 + abr - 2abr)/a^2b^2r^2$$

$$= \sigma_r^2(r-1)/r + \sigma_d^2(r-1)/ar + \sigma_e^2(r-1)/abr$$

Thus, $E[ab\sum (\bar{y}_{i..} - \bar{y}_{...})^2/(r-1)] = \sigma_e^2 + b\sigma_d^2 + ab\sigma_r^2$

As can be seen from the expected mean squares, the comparisons between factor A effects is a between whole plot experimental unit comparison. Thus, the proper F-statistic for comparing equal factor A effects is MSA/MSE(a). The comparisons between factor B effects and the interaction of factors A and B are between sub-plot experimental unit comparisons. Thus, the proper F-statistics for comparing equal factor B effects and for comparing the interactions are MSB/MSE(b) and MSAB/MSE(b), respectively.

Moreover, if the allocation of the whole plot treatments is done either in completely random fashion or Latin Square arrangement, then it can be shown that the expected mean squares for the source of variation due to the main effects and the interaction effects are exactly the same. Likewise, the F-statistic to test the significance of the main effects and interaction effects is also the same. Table 2 shows the various expected mean squares for experiments laid out in split-plot design where the whole plot treatments are arranged in a completely randomized design.

Table 2. Analysis of Variance Table for Split-Plot Design
(whole plot units arranged in CRD)

Source	df	SS	MS	E(MS)
A	a-1	SSA	MSA	$\sigma_e^2 + b\sigma_d^2 + br\Sigma\, \alpha_i^2/(a-1)$
Error(a)	a(r-1)	SSE(a)	MSE(a)	$\sigma_e^2 + b\sigma_d^2$
B	b-1	SSB	MSB	$\sigma_e^2 + ar\Sigma\, \beta_j^2/(b-1)$
A*B	(a-1)(b-1)	SSAB	MSAB	$\sigma_e^2 + r\Sigma\Sigma\, (\alpha\beta)_{ij}^2/(a-1)(b-1)$
Error(b)	a(b-1)(r-1)	SSE(b)	MSE(b)	σ_e^2
Total	abr-1	SST		

For a split plot design where the whole plot treatments are arranged in Latin Square design, the analysis of variance is shown in Table 3.

Table 3. Analysis of Variance Table for a Split Plot Design (whole plot units arranged in Latin Square)

Source	df	SS	MS	E(MS)
Rows	a-1	SSR	MSR	$\sigma_e^2 + b\sigma_d^2 + ab\sigma_r^2$
Columns	a-1	SSC	MSC	$\sigma_e^2 + b\sigma_d^2 + ab\sigma_c^2$
A	a-1	SSA	MSA	$\sigma_e^2 + b\sigma_d^2 + ab\Sigma\, \alpha_k^2/(a-1)$
Error(a)	(a-1)(a-2)	SSE(a)	MSE(a)	$\sigma_e^2 + b\sigma_d^2$
B	b-1	SSB	MSB	$\sigma_e^2 + a^2\Sigma\, \beta_1^2/(b-1)$
A*B	(a-1)(b-1)	SSAB	MSAB	$\sigma_e^2 + a\Sigma\Sigma\, (\alpha\beta)_{k1}^2/(a-1)(b-1)$
Error(b)	a(a-1)(b-1)	SSE(b)	MSE(b)	σ_e^2
Total	a^2b-1	SST		

(4) Multiple Comparisons.

Once the F-tests have been made to determine if there are significant differences between means, the next step is to carry out multiple comparisons to determine where the differences occur. The usual least significant differences test (LSD) can be employed. Hence, results will be derived for finding the appropriate standard error to use.

To compare the difference between two A means, that is,

$$Ho: \bar{\mu}_{j.} - \bar{\mu}_{j'.} = 0 \quad \text{versus} \quad Ha: \bar{\mu}_{j.} - \bar{\mu}_{j'.} \neq 0 \quad \text{for } j \neq j' ,$$ we

determine the corresponding variance.

$$V(\bar{y}_{.j.} - \bar{y}_{.j'.}) = V(\bar{\mu}_{j.} + \bar{\rho}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.j.} - \bar{\mu}_{j'.} - \bar{\rho}_{.} - \bar{\delta}_{.j'} - \bar{\varepsilon}_{.j'.})$$

$$= V(\bar{\delta}_{.j} - \bar{\delta}_{.j'} + \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{.j'.})$$

$$= V(\bar{\delta}_{.j} - \bar{\delta}_{.j'}) + V(\bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{.j'.})$$

$$= V(\delta_{ij} - \delta_{ij'})/r + V(\varepsilon_{ijk} - \varepsilon_{ij'k})/br$$

$$= 2\sigma_d^2/r + 2\sigma_e^2/rb$$

$$= 2(b\sigma_d^2 + \sigma_e^2)/rb$$

It follows from the table of expected mean squares that

$$\widehat{V(\bar{y}_{.j.} - \bar{y}_{.j'.})} = 2\ \mathrm{MSE(a)}/rb$$

To compare the difference between two B means, that is,

$$\mathrm{Ho}: \bar{\mu}_{.k} - \bar{\mu}_{.k'} = 0 \quad \text{versus} \quad \mathrm{Ha}: \bar{\mu}_{.k} - \bar{\mu}_{.k'} \neq 0 \quad \text{for } k \neq k' \text{, we}$$

determine the corresponding variance.

$$V(\bar{y}_{..k} - \bar{y}_{..k'}) = V(\bar{\mu}_{.k} + \bar{\rho}_{.} + \bar{\delta}_{..} + \bar{\varepsilon}_{..k} - \bar{\mu}_{.k'} - \bar{\rho}_{.} - \bar{\delta}_{..} - \bar{\varepsilon}_{..k'})$$

$$= V(\bar{\varepsilon}_{..k} - \bar{\varepsilon}_{..k'})$$

$$= V(\varepsilon_{ijk} - \varepsilon_{ijk'})/ar$$

$$= 2\sigma_e^2/ar$$

It follows from the table of expected mean squares that

$$\widehat{V(\bar{y}_{..k} - \bar{y}_{..k'})} = 2\ \mathrm{MSE(b)}/ar$$

When there is a significant A*B interaction, comparisons must be based on the set of two-way cell means. There are two different types of comparison one must consider when comparing the 2 cell means. The first type arises when two sub-plot treatment means are compared at the same level of a whole plot treatment. That is,

$$Ho: \mu_{jk} - \mu_{jk'} = 0 \quad \text{versus} \quad Ha: \mu_{jk} - \mu_{jk'} \neq 0 \quad \text{for all j and } k \neq k'$$

The corresponding variance is

$$V(\bar{y}_{.jk} - \bar{y}_{.jk'}) = V[\mu_{jk} + \bar{\rho}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.jk} - \mu_{jk'} - \bar{\rho}_{.} - \bar{\delta}_{.j} - \bar{\varepsilon}_{.jk'}]$$

$$= V(\bar{\varepsilon}_{.jk} - \bar{\varepsilon}_{.jk'})$$

$$= V(\varepsilon_{ijk} - \varepsilon_{ijk'})/r$$

$$= 2\sigma_e^2/r$$

Thus, the estimate of this variance is

$$\widehat{V(\bar{y}_{.jk} - \bar{y}_{.jk'})} = 2\ \mathrm{MSE(b)}/r$$

The second type of comparison occurs when the two whole plot treatments are compared at the same level or different levels of the sub-plot treatments. That is,

$$Ho: \mu_{jk} - \mu_{j'k} = 0 \quad \text{or} \quad Ho: \mu_{jk} - \mu_{j'k'} = 0 \quad \text{versus}$$

$$Ha: \mu_{jk} - \mu_{j'k} \neq 0 \quad \text{or} \quad Ha: \mu_{jk} - \mu_{j'k'} \neq 0$$

$$\text{for all k and } k' \text{ and } j \neq j'$$

The corresponding variance for this comparison (which turns out to be the same for the two types of hypothesis) is

$$V(\bar{y}_{.jk} - \bar{y}_{.j'k}) = V[\mu_{jk} + \bar{\rho}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.jk} - \mu_{j'k} - \bar{\rho}_{.} - \bar{\delta}_{.j'} - \bar{\varepsilon}_{.j'k}]$$

$$= V(\bar{\delta}_{.j} - \bar{\delta}_{.j'} + \bar{\varepsilon}_{.jk} - \bar{\varepsilon}_{.j'k})$$

$$= V(\bar{\delta}_{.j} - \bar{\delta}_{.j'}) + V(\bar{\varepsilon}_{.jk} - \bar{\varepsilon}_{.j'k})$$

$$= V(\delta_{ij} - \delta_{ij'})/r + V(\varepsilon_{ijk} - \varepsilon_{ij'k})/r$$

$$= 2\sigma_d^2/r + 2\sigma_e^2/r$$

$$= 2(\sigma_d^2 + \sigma_e^2)/r$$

Thus, the estimate of this variance is

$$\widehat{V(\bar{y}_{.jk} - \bar{y}_{.j'k})} = 2[MSE(a) + (b-1)MSE(b)]/br$$

Given the estimates of the variance for each effects, the LSD can be computed in the usual manner, that is,

$$LSD_\alpha = t_{\alpha/2,\ \text{error df}}(SE).$$

For each comparisons, the LSD can be written as

(a) difference between two A means

$$LSD_\alpha = t_{\alpha/2,\ (r-1)(a-1)} \sqrt{2MSE(a)/br}$$

(b) difference between two B means

$$LSD_\alpha = t_{\alpha/2,\ a(b-1)(r-1)} \sqrt{2MSE(b)/ar}$$

(c) difference between two A*B means

i) different levels of B with the same level of A

$$LSD_\alpha = t_{\alpha/2,\ a(b-1)(r-1)} \sqrt{2MSE(b)/r}$$

ii) different levels of A with the same or different levels of B

$$LSD_\alpha = t^* \sqrt{2[MSE(a) + (b-1)MSE(b)]/br}$$

where $$t^* = \frac{[t_{\alpha/2,(a-1)(r-1)}\mathrm{MSE}(a) + t_{\alpha/2,a(b-1)(r-1)}(b-1)\mathrm{MSE}(b)]}{\mathrm{MSE}(a) + (b-1)\mathrm{MSE}(b)}$$

is an approximation to correct the degrees of freedom furnished by Satterthwaite (21).

III. VARIATIONS OF SPLIT-PLOT DESIGN

(1) Two factors whole plot and single factor sub-plot.

Suppose the experiment is originally planned to test the effects of 2 factors, say factor A and factor B, as well as their interaction. The experiment is then laid out using a Randomized Complete Block arrangement where each plot receives a combination of factor A and B. The division of each plot into several sub-plot permits the inclusion of an extra factor C. Thus within each level of the treatment combination A*B, the levels of C are allocated at random. The plan after randomization might appear as follows:

	Replication I			Replication II			Replication III	
a_1b_3	c_1	c_2	a_1b_1	c_2	c_1	a_2b_1	c_1	c_2
a_1b_1	c_2	c_1	a_1b_2	c_1	c_2	a_1b_2	c_2	c_1
a_2b_3	c_1	c_2	a_2b_1	c_2	c_1	a_1b_3	c_2	c_1
a_2b_1	c_2	c_1	a_2b_3	c_1	c_2	a_2b_3	c_1	c_2
a_1b_2	c_1	c_2	a_2b_2	c_2	c_1	a_1b_1	c_2	c_1
a_2b_2	c_1	c_2	a_1b_3	c_2	c_1	a_2b_2	c_1	c_2

where factor A consists of 2 levels, factor B with 3 levels, factor C with 2 levels and three replications.

The model to describe this type of experiment is also a split-plot design where the main plot treatments consists of the combinations of factors A and B and the sub-plot treatments are the levels of factor C. The mathematical model can be written as

$$y_{ijkl} = \mu + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \delta_{ijk} + \gamma_l + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl} + \varepsilon_{ijkl} \qquad (4)$$

$i=1,2,\ldots,r \quad j=1,2,\ldots,a \quad k=1,2,\ldots,b \quad l=1,2,\ldots,c$

where y_{ijkl} = an individual observation taken on the i<u>th</u> block of the l<u>th</u> level of factor C on the jk<u>th</u> level of factor A and B.

μ = overall mean

ρ_i = block effect

α_j = factor A effect

β_k = factor B effect

$(\alpha\beta)_{jk}$ = interaction of factor A and B effect

δ_{ijk} = residual main plot effect [error(ab)]

γ_l = sub-plot treatment effect

$(\alpha\gamma)_{jl}$ = interaction of factor A and C effect

$(\beta\gamma)_{kl}$ = interaction of factor B and C effect

$(\alpha\beta\gamma)_{jkl}$ = interaction of factors A,B and C effect

ε_{ijkl} = random error sub-plot [error(c)]

To carry out the analysis for this experiment, the usual assumptions hold:

1) $\delta_{ijk} \sim N(0,\sigma_d^2)$

2) $\varepsilon_{ijkl} \sim N(0,\sigma_e^2)$

3) δ_{ijk} and ε_{ijkl} are independent

4) Factors A, B and C are fixed

5) Blocks are random

6) $\sum_j \alpha_j = \sum_k \beta_k = \sum_l \gamma_l = \sum_j (\alpha\beta)_{jk} = \sum_k (\alpha\beta)_{jk} = \sum_j (\alpha\gamma)_{jl} = \sum_l (\alpha\gamma)_{jl}$

$= \sum_k (\beta\gamma)_{kl} = \sum_l (\beta\gamma)_{kl} = \sum_j (\alpha\beta\gamma)_{jkl} = \sum_k (\alpha\beta\gamma)_{jkl} = \sum_l (\alpha\beta\gamma)_{jkl} = 0$

The analysis of variance is given in Table 4.

Table 4. Analysis of Variance Table for a Split Plot Design (two whole plot factor and single sub-plot factor)

Source	df	SS	MS	E(MS)
Replication	r-1	SSR	MSR	$\sigma_e^2 + c\sigma_d^2 + abc\sigma_r^2$
A	a-1	SSA	MSA	$\sigma_e^2 + c\sigma_d^2 + bcr\Sigma\, \alpha_j^2/(a-1)$
B	b-1	SSB	MSB	$\sigma_e^2 + c\sigma_d^2 + acr\Sigma\, \beta_k^2/(b-1)$
A*B	(a-1)(b-1)	SSAB	MSAB	$\sigma_e^2 + c\sigma_d^2 + cr\Sigma\Sigma\, (\alpha\beta)_{jk}^2/(a-1)(b-1)$
Error(ab)	(ab-1)(r-1)	SSE(ab)	MSE(ab)	$\sigma_e^2 + c\sigma_d^2$
C	c-1	SSC	MSC	$\sigma_e^2 + abr\Sigma\, \gamma_1^2/(c-1)$
A*C	(a-1)(c-1)	SSAC	MSAC	$\sigma_e^2 + br\Sigma\Sigma\, (\alpha\gamma)_{j1}^2/(a-1)(c-1)$
B*C	(b-1)(c-1)	SSBC	MSBC	$\sigma_e^2 + ar\Sigma\Sigma\, (\beta\gamma)_{k1}^2/(b-1)(c-1)$
A*B*C	(a-1)(b-1)(c-1)	SSABC	MSABC	$\sigma_e^2 + r\Sigma\Sigma\Sigma\, (\alpha\beta\gamma)_{jk1}^2/(a-1)(b-1)(c-1)$
Error(c)	ab(c-1)(r-1)	SSE(c)	MSE(c)	σ_e^2
Total	abcr-1	SST		

A lengthy but straightforward derivation of the expected mean squares can be obtained by following the procedure of the classical split-plot design.

The variance for the various comparisons are:

a) Between two A means

$$Ho: \bar{\mu}_{j..} - \bar{\mu}_{j'..} = 0 \quad \text{versus} \quad Ha: \bar{\mu}_{j..} - \bar{\mu}_{j'..} \neq 0 \qquad \text{for } j \neq j'$$

$$V(\bar{y}_{.j..} - \bar{y}_{.j'..}) = V(\bar{\mu}_{j..} + \bar{\rho}_{.} + \bar{\delta}_{.j.} + \bar{\varepsilon}_{.j..} - \bar{\mu}_{j'..} - \bar{\rho}_{.} - \bar{\delta}_{.j'.} - \bar{\varepsilon}_{.j'..})$$

$$= V(\bar{\delta}_{.j.} - \bar{\delta}_{.j'.}) + V(\bar{\varepsilon}_{.j..} - \bar{\varepsilon}_{.j'..})$$

$$= 2\sigma_d^2/br + 2\sigma_e^2/bcr$$

$$= 2(\sigma_e^2 + c\sigma_d^2)/bcr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.j..} - \bar{y}_{.j'..})} = 2MSE(ab)/bcr$$

b) Between two B means

$$Ho: \bar{\mu}_{.k.} - \bar{\mu}_{.k'.} = 0 \quad \text{versus} \quad Ha: \bar{\mu}_{.k.} - \bar{\mu}_{.k'.} \neq 0 \qquad \text{for } k \neq k'$$

$$V(\bar{y}_{..k.} - \bar{y}_{..k'.}) = V(\bar{\mu}_{.j.} + \bar{\rho}_{.} + \bar{\delta}_{..k} + \bar{\varepsilon}_{..k.} - \bar{\mu}_{.j'.} - \bar{\rho}_{k'} - \bar{\delta}_{..k'} - \bar{\varepsilon}_{..k'.})$$

$$= V(\bar{\delta}_{..k} - \bar{\delta}_{..k'}) + V(\bar{\varepsilon}_{..k.} - \bar{\varepsilon}_{..k'.})$$

$$= 2\sigma_d^2\ /ar + 2\sigma_e^2/acr$$

$$= 2(\sigma_e^2 + c\sigma_d^2)/acr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{..k.} - \bar{y}_{..k'.})} = 2MSE(ab)/acr$$

c) Between two A*B means

$$\text{Ho: } \bar{\mu}_{jk.} - \bar{\mu}_{j'k'.} = 0 \quad \text{versus} \quad \text{Ha: } \bar{\mu}_{jk.} - \bar{\mu}_{j'k'.} \neq 0 \quad \text{for } j \neq j' \text{ and } k \neq k'$$

$$V(\bar{y}_{.jk.} - \bar{y}_{.j'k'.}) = V(\bar{\mu}_{jk.} + \bar{\rho}_{.} + \bar{\delta}_{.jk} + \bar{\varepsilon}_{.jk.} - \bar{\mu}_{j'k'.} - \bar{\rho}_{.} - \bar{\delta}_{.j'k'} - \bar{\varepsilon}_{.j'k'.})$$

$$= V(\bar{\delta}_{.jk} - \bar{\delta}_{.j'k'}) + V(\bar{\varepsilon}_{.jk.} - \bar{\varepsilon}_{.j'k'.})$$

$$= 2\sigma_d^2/r + 2\sigma_e^2/cr$$

$$= 2(\sigma_e^2 + c\sigma_d^2)/cr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.jk.} - \bar{y}_{.j'k'.})} = 2MSE(ab)/cr$$

d) Between two C means

$$\text{Ho: } \bar{\mu}_{..l} - \bar{\mu}_{..l'} = 0 \quad \text{versus} \quad \text{Ha: } \bar{\mu}_{..l} - \bar{\mu}_{..l'} \neq 0 \quad \text{for } l \neq l'$$

$$V(\bar{y}_{...l} - \bar{y}_{...l'}) = V(\bar{\mu}_{..l} + \bar{\rho}_{.} + \bar{\delta}_{...} + \bar{\varepsilon}_{...l} - \bar{\mu}_{..l'} - \bar{\rho}_{.} - \bar{\delta}_{...} - \bar{\varepsilon}_{...l'})$$

$$= V(\bar{\varepsilon}_{...l} - \bar{\varepsilon}_{...l'})$$

$$= 2\sigma_e^2/abr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{...l} - \bar{y}_{...l'})} = 2MSE(c)/abr$$

e) Between two A*C means

i) same level of A with different levels of C

Ho: $\bar{\mu}_{j.l} - \bar{\mu}_{j.l'} = 0$ versus Ha: $\bar{\mu}_{j.l} - \bar{\mu}_{j.l'} = 0$ for all j and $l \neq l'$

$$V(\bar{y}_{.j.l} - \bar{y}_{.j.l'}) = V(\bar{\mu}_{j.l} + \bar{\rho}_{.} + \bar{\delta}_{.j.} + \bar{\varepsilon}_{.j.l} - \bar{\mu}_{j.l'} - \bar{\rho}_{.} - \bar{\delta}_{.j.} - \bar{\varepsilon}_{.j.l'})$$

$$= V(\bar{\varepsilon}_{.j.l} - \bar{\varepsilon}_{.j.l'})$$

$$= 2\sigma_e^2/br$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.j.l} - \bar{y}_{.j.l'})} = 2MSE(c)/br$$

ii) different A levels with different (or same) levels of C

Ho: $\bar{\mu}_{j.l} - \bar{\mu}_{j'.l} = 0$ or Ho: $\bar{\mu}_{j.l} - \bar{\mu}_{j'.l'} = 0$ versus

Ha: $\bar{\mu}_{j.l} - \bar{\mu}_{j'.l} \neq 0$ or Ha: $\bar{\mu}_{j.l} - \bar{\mu}_{j'.l'} \neq 0$ for all l, l' and $j \neq j'$

$$V(\bar{y}_{.j.l} - \bar{y}_{.j'.l}) = V(\bar{\mu}_{j.l} + \bar{\rho}_{.} + \bar{\delta}_{.j.} + \bar{\varepsilon}_{.jl.} - \bar{\mu}_{j'.l} - \bar{\rho}_{.} - \bar{\delta}_{.j'.} - \bar{\varepsilon}_{.j'.l})$$

$$= V(\bar{\delta}_{.j.} - \bar{\delta}_{.j'.}) + V(\bar{\varepsilon}_{.j.l} - \bar{\varepsilon}_{.j'.l})$$

$$= 2\sigma_d^2/br + 2\sigma_e^2/br$$

$$= 2(\sigma_d^2 + \sigma_e^2)/br$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.j.l} - \bar{y}_{.j'.l})} = 2[MSE(ab) + (c-1)MSE(c)]/bcr$$

f) Between two B*C means

i) same level of B with different C levels

$$Ho: \bar{\mu}_{.kl} - \bar{\mu}_{.kl'} = 0 \quad \text{versus } Ha: \bar{\mu}_{.kl} - \bar{\mu}_{.kl'} \neq 0 \quad \text{for } l \neq l'$$

$$V(\bar{y}_{..kl} - \bar{y}_{..kl'}) = V(\bar{\mu}_{.kl} + \bar{\rho}_{.} + \bar{\delta}_{..k} + \bar{\varepsilon}_{..kl} - \bar{\mu}_{.kl'} - \bar{\rho}_{.} - \bar{\delta}_{..k} - \bar{\varepsilon}_{..kl'})$$

$$= V(\bar{\varepsilon}_{..kl} - \bar{\varepsilon}_{..kl'})$$

$$= 2\sigma_e^2/ar$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{..kl} - \bar{y}_{..kl'})} = 2MSE(c)/ar$$

ii) different levels of B with different (or same) levels of C

$$Ho: \bar{\mu}_{.kl} - \bar{\mu}_{.k'l} = 0 \quad \text{or } Ho: \bar{\mu}_{.kl} - \bar{\mu}_{.k'l'} = 0 \quad \text{versus}$$

$$Ha: \bar{\mu}_{.kl} - \bar{\mu}_{.k'l} \neq 0 \quad \text{or } Ha: \bar{\mu}_{.kl} - \bar{\mu}_{.k'l'} \neq 0 \quad \text{for all } l,\ l' \text{ and } k \neq k'$$

$$V(\bar{y}_{..kl} - \bar{y}_{..k'l}) = V(\bar{\mu}_{.kl} + \bar{\rho}_{.} + \bar{\delta}_{..k} + \bar{\varepsilon}_{..kl} - \bar{\mu}_{.k'l} - \bar{\rho}_{.} - \bar{\delta}_{..k'} - \bar{\varepsilon}_{..k'l})$$

$$= V(\bar{\delta}_{..k} - \bar{\delta}_{..k'}) + V(\bar{\varepsilon}_{..kl} - \bar{\varepsilon}_{..k'l})$$

$$= 2\sigma_d^2/ar + 2\sigma_e^2/ra$$

$$= 2(\sigma_d^2 + \sigma_e^2)/ar$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{..kl} - \bar{y}_{..k'l})} = 2[MSE(ab) + (c-1)MSE(c)]/acr$$

(g) Between two A*B*C means

i) same level of A*B with different C levels

Ho: $\mu_{jkl} - \mu_{jkl'} = 0$ versus Ha: $\mu_{jkl} - \mu_{jkl'} \neq 0$ for $l \neq l'$

$$V(\bar{y}_{.jkl} - \bar{y}_{.jkl'}) = V(\mu_{jkl} + \bar{\rho}_{.} + \bar{\delta}_{.jk} + \bar{\varepsilon}_{.jkl} - \mu_{jkl'} - \bar{\rho}_{.} - \bar{\delta}_{.jk} - \bar{\varepsilon}_{.jkl'})$$

$$= V(\bar{\varepsilon}_{.jkl} - \bar{\varepsilon}_{.jkl'})$$

$$= 2\sigma_e^2/r$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.jkl} - \bar{y}_{.jkl'})} = 2MSE(c)/r$$

ii) different A*B levels with different (or same) C levels

Ho: $\mu_{jkl} - \mu_{j'k'l} = 0$ or Ho: $\mu_{jkl} - \mu_{j'k'l'} = 0$ vs

Ha: $\mu_{jkl} - \mu_{j'k'l} \neq 0$ or Ha: $\mu_{jkl} - \mu_{j'k'l'} \neq 0$

for all l and l' and $j \neq j'$ and $k \neq k'$

$$V(\bar{y}_{.jkl} - \bar{y}_{.j'k'l}) = V(\mu_{jkl} + \bar{\rho}_{.} + \bar{\delta}_{.jk} + \bar{\varepsilon}_{.jkl} - \mu_{j'k'l} - \bar{\rho}_{.} - \bar{\delta}_{.j'k'} - \bar{\varepsilon}_{.j'k'l})$$

$$= V(\bar{\delta}_{.jk} - \bar{\delta}_{.j'k'}) + V(\bar{\varepsilon}_{.jkl} - \bar{\varepsilon}_{.j'k'l})$$

$$= 2\sigma_d^2/r + 2\sigma_e^2/r$$

$$= 2(\sigma_d^2 + \sigma_e^2)/r$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.jkl} - \bar{y}_{.j'k'l})} = 2[MSE(ab) + (c-1)MSE(c)]/rc$$

The LSD value can be obtained by using the above estimates of the variance, that is,

a) difference between two A means

$$LSD_\alpha = t_{\alpha/2,(ab-1)(r-1)} \sqrt{2MSE(ab)/bcr}$$

b) difference between two B means

$$LSD_\alpha = t_{\alpha/2,(ab-1)(r-1)} \sqrt{2MSE(ab)/acr}$$

c) difference between two C means

$$LSD_\alpha = t_{\alpha/2,ab(c-1)(r-1)} \sqrt{2MSE(c)/abr}$$

d) difference between two A*B means

$$LSD_\alpha = t_{\alpha/2,(ab-1)(r-1)} \sqrt{2MSE(ab)/cr}$$

e) difference between two A*C means

i) different levels of C with same level of A

$$LSD_\alpha = t_{\alpha/2,ab(c-1)(r-1)} \sqrt{2MSE(c)/br}$$

ii) different levels of A with different (or same) level of C

$$LSD_\alpha = t^* \sqrt{2[MSE(ab) + (c-1)MSE(c)]/bcr}$$

f) difference between 2 B*C means

i) different levels of C with same level of B

$$LSD_\alpha = t_{\alpha/2,ab(c-1)(r-1)} \sqrt{2MSE(c)/ar}$$

ii) different levels of B with different (or same) level of C

$$LSD_\alpha = t^* \sqrt{2[MSE(ab) + (c-1)MSE(c)]/acr}$$

g) difference between 2 A*B*C means

i) different levels of C with same level of A*B

$$LSD_{\alpha} = t_{\alpha/2,ab(c-1)(r-1)} \sqrt{2MSE(c)/r}$$

ii) different levels of A*B with different (or same) level of C

$$LSD_{\alpha} = t^{*} \sqrt{2[MSE(ab) + (c-1)MSE(c)]/cr}$$

where $$t^{*} = \frac{t_{\alpha/2,(ab-1)(r-1)}MSE(ab) + t_{\alpha/2,ab(c-1)(r-1)}(c-1)MSE(c)}{MSE(ab) + (c-1)MSE(c)}$$

(2) Single factor whole plot and 2 factors sub-plot.

Suppose we plan the experiment where the whole plot treatments consist of a levels of factor A and the sub-plot treatments consist of the b*c levels of factors B and C. The model of this experiment is

$$y_{ijkl} = \mu + \rho_i + \alpha_j + \delta_{ij} + \beta_k + \gamma_l + (\alpha\beta)_{jk} + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl} + \varepsilon_{ijkl} \qquad (5)$$

$$i=1,2,\ldots,r \quad j=1,2,\ldots,a \quad k=1,2,\ldots,b \quad l=1,2,\ldots,c$$

where y_{ijkl} = an observation on the ith block of the klth sub-plot on the jth plot.

μ = overall mean

ρ_i = block effect

α_j = main plot treatment effect

δ_{ij} = main plot residual effect [error(a)]

β_k = factor B effect

γ_l = factor C effect

$(\alpha\beta)_{jk}$ = interaction of A*B effect

$(\alpha\gamma)_{jl}$ = interaction of A*C effect

$(\beta\gamma)_{kl}$ = interaction of B*C effect

$(\alpha\beta\gamma)_{jkl}$ = interaction of A*B*C effect

ε_{ijkl} = random sub-plot error [error(bc)]

The analysis of variance table for this field arrangement is given below.

Table 5. Analysis of Variance Table for a Split Plot Design (single factor whole plot and two factors sub-plot)

Source	df	SS	MS	E(MS)
Replication	r-1	SSR	MSR	$\sigma_e^2 + bc\sigma_d^2 + abc\sigma_r^2$
A	a-1	SSA	MSA	$\sigma_e^2 + bc\sigma_d^2 + bcr\Sigma\,\alpha_j^2/(a-1)$
Error(a)	(a-1)(b-1)	SSE(a)	MSE(a)	$\sigma_e^2 + bc\sigma_d^2$
B	b-1	SSB	MSB	$\sigma_e^2 + acr\Sigma\,\beta_k^2/(b-1)$
C	c-1	SSC	MSC	$\sigma_e^2 + abr\Sigma\,\gamma_1^2/(c-1)$
B*C	(b-1)(c-1)	SSBC	MSBC	$\sigma_e^2 + ar\Sigma\Sigma\,(\beta\gamma)_{k1}^2/(b-1)(c-1)$
A*B	(a-1)(b-1)	SSAB	MSAB	$\sigma_e^2 + cr\Sigma\Sigma\,(\alpha\beta)_{jk}^2/(a-1)(b-1)$
A*C	(a-1)(b-1)	SSAC	MSAC	$\sigma_e^2 + br\Sigma\Sigma\,(\alpha\gamma)_{j1}^2/(a-1)(c-1)$
A*B*C	(a-1)(b-1)(c-1)	SSABC	MSABC	$\sigma_e^2 + r\Sigma\Sigma\Sigma\,(\alpha\beta\gamma)_{jk1}^2/(a-1)(b-1)(c-1)$
Error(bc)	a(bc-1)(r-1)	SSE(bc)	MSE(bc)	σ_e^2
Total	abcr-1	SST		

The estimates of the variance for the various comparisons are:

a) Difference between two A means

$$\text{Ho: } \bar{\mu}_{j..} - \bar{\mu}_{j'..} = 0 \quad \text{versus Ha: } \bar{\mu}_{j..} - \bar{\mu}_{j'..} \neq 0 \qquad \text{for } j \neq j'$$

$$V(\bar{y}_{.j..} - \bar{y}_{.j'..}) = V(\bar{\mu}_{j..} + \bar{\rho}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.j..} - \bar{\mu}_{j'..} - \bar{\rho}_{.} - \bar{\delta}_{.j'} - \bar{\varepsilon}_{.j'..})$$

$$= V(\bar{\delta}_{.j} - \bar{\delta}_{.j'}) + V(\bar{\varepsilon}_{.j..} - \bar{\varepsilon}_{.j'..})$$

$$= 2\sigma_d^2/r + 2\sigma_e^2/bcr$$

$$= 2(\sigma_e^2 + bc\sigma_d^2)/bcr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.j..} - \bar{y}_{.j'..})} = 2MSE(a)/bcr$$

b) Difference between two B means

Ho: $\bar{\mu}_{.k.} - \bar{\mu}_{.k'.} = 0$ versus Ha: $\bar{\mu}_{.k.} - \bar{\mu}_{.k'.} \neq 0$ for $k \neq k'$

$$V(\bar{y}_{..k.} - \bar{y}_{..k'.}) = V(\bar{\mu}_{.k.} + \bar{\rho}_{.} + \bar{\delta}_{..} + \bar{\varepsilon}_{..k.} - \bar{\mu}_{.k'.} - \bar{\rho}_{.} - \bar{\delta}_{..} - \bar{\varepsilon}_{..k'.})$$

$$= V(\bar{\varepsilon}_{..k.} - \bar{\varepsilon}_{..k'.})$$

$$= 2\sigma_e^2/acr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{..k.} - \bar{y}_{..k'.})} = 2MSE(bc)/acr$$

c) Difference between two C means

Ho: $\mu_{..1} - \mu_{..1'} = 0$ versus Ha: $\mu_{..1} - \mu_{..1'} \neq 0$ for $1 \neq 1'$

$$V(\bar{y}_{...1} - \bar{y}_{...1'}) = V(\bar{\mu}_{..1} + \bar{\rho}_{.} + \bar{\delta}_{..} + \bar{\varepsilon}_{...1} - \bar{\mu}_{..1'} - \bar{\rho}_{.} - \bar{\delta}_{..} - \bar{\varepsilon}_{...1'})$$

$$= V(\bar{\varepsilon}_{...1} - \bar{\varepsilon}_{...1'})$$

$$= 2\sigma_e^2/abr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{...l} - \bar{y}_{...l'})} = 2MSE(bc)/abr$$

d) Difference between two A*B means

i) same level of A with different levels of B

Ho: $\bar{\mu}_{jk.} - \bar{\mu}_{jk'.} = 0$ versus Ha: $\bar{\mu}_{jk.} - \bar{\mu}_{jk'.} \neq 0$ for all j and k ≠ k'

$$V(\bar{y}_{.jk.} - \bar{y}_{.jk'.}) = V(\bar{\mu}_{jk.} + \bar{\rho}_{.} + \bar{\delta}_{.j} + + \bar{\varepsilon}_{.jk.} - \bar{\mu}_{jk'.} - \bar{\rho}_{.} - \bar{\delta}_{.j} - \bar{\varepsilon}_{.jk'.})$$

$$= V(\bar{\varepsilon}_{.jk.} - \bar{\varepsilon}_{.jk'.})$$

$$= 2\sigma_e^2/cr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.jk.} - \bar{y}_{.jk'.})} = 2MSE(bc)/cr$$

ii) different levels of A with different (or same) level of B

Ho: $\bar{\mu}_{jk.} - \bar{\mu}_{j'k.} = 0$ or Ho: $\bar{\mu}_{jk.} - \bar{\mu}_{j'k'.} = 0$ versus

Ha: $\bar{\mu}_{jk.} - \bar{\mu}_{j'k.} \neq 0$ or Ha: $\bar{\mu}_{jk.} - \bar{\mu}_{j'k'.} \neq 0$ for all k, k' and j ≠ j'

$$V(\bar{y}_{.jk.} - \bar{y}_{.j'k.}) = V(\bar{\mu}_{jk.} + \bar{\rho}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.jk.} - \bar{\mu}_{j'k.} - \bar{\rho}_{.} - \bar{\delta}_{.j'} - \bar{\varepsilon}_{.j'k.})$$

$$= V(\bar{\delta}_{.j} - \bar{\delta}_{.j'}) + V(\bar{\varepsilon}_{.jk.} - \bar{\varepsilon}_{.j'k.})$$

$$= 2\sigma_d^2/r + 2\sigma_e^2/cr$$

$$= 2(\sigma_e^2 + c\sigma_d^2)/cr$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.jk.} - \bar{y}_{.j'k.})} = 2[MSE(a) + (b-1)MSE(bc)]/bcr$$

e) Difference between two A*C means

i) same level of A with different levels of C

Ho: $\bar{\mu}_{j.l} - \bar{\mu}_{j.l'} = 0$ versus Ha: $\bar{\mu}_{j.l} - \bar{\mu}_{j.l'} \neq 0$ for $l \neq l'$

$$V(\bar{y}_{.j.l} - \bar{y}_{.j.l'}) = V(\bar{\mu}_{j.l} + \bar{\rho}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.j.l} - \bar{\mu}_{j.l'} - \bar{\rho}_{.} - \bar{\delta}_{.j} - \bar{\varepsilon}_{.j.l'})$$

$$= V(\bar{\varepsilon}_{.j.l} - \bar{\varepsilon}_{.j.l'})$$

$$= 2\sigma_e^2/br$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.j.l} - \bar{y}_{.j.l'})} = 2MSE(bc)/br$$

ii) different levels of A with different (or same) level of C

Ho: $\bar{\mu}_{j.l} - \bar{\mu}_{j'.l} = 0$ or Ho: $\bar{\mu}_{j.l} - \bar{\mu}_{j'.l'} = 0$ versus

Ha: $\bar{\mu}_{j.l} - \bar{\mu}_{j'.l} \neq 0$ or Ha: $\bar{\mu}_{j.l} - \bar{\mu}_{j'.l'} \neq 0$ for all l and l' and $j \neq j'$

$$V(\bar{y}_{.j.l} - \bar{y}_{.j'.l}) = V(\bar{\mu}_{j.l} + \bar{p}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.j.l} - \bar{\mu}_{j'.l} - \bar{p}_{.} - \bar{\delta}_{.j'} - \bar{\varepsilon}_{.j'.l})$$

$$= V(\bar{\delta}_{.j} - \bar{\delta}_{.j'}) + V(\bar{\varepsilon}_{.j.l} - \bar{\varepsilon}_{.j'.l})$$

$$= 2\sigma_d^2/r + 2\sigma_e^2/br$$

$$= 2(\sigma_e^2 + b\sigma_d^2)/br$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.j.l} - \bar{y}_{.j'.l})} = 2[MSE(a) + (c-1)MSE(bc)]/bcr$$

f) Difference between two B*C means

Ho: $\bar{\mu}_{.kl} - \bar{\mu}_{.k'l'} = 0$ versus Ha: $\bar{\mu}_{.kl} - \bar{\mu}_{.k'l'} \neq 0$ for $k \neq k'$, $l \neq l'$

$$V(\bar{y}_{..kl} - \bar{y}_{..k'l'}) = V(\bar{\mu}_{.kl} + \bar{\rho}_{.} + \bar{\delta}_{..} + \bar{\varepsilon}_{..kl} - \bar{\mu}_{.k'l'} - \bar{\rho}_{.} - \bar{\delta}_{..} - \bar{\varepsilon}_{..k'l'})$$

$$= V(\bar{\varepsilon}_{..kl} - \bar{\varepsilon}_{..k'l'})$$

$$= 2\sigma_e^2/ar$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{..kl} - \bar{y}_{..k'l'})} = 2MSE(bc)/ar$$

g) Difference between two A*B*C means

i) same level of A with different levels of B*C

Ho: $\mu_{jkl} - \mu_{jk'l'} = 0$ or Ho: $\mu_{jkl} - \mu_{jk'l} = 0$ or Ho: $\mu_{jkl} - \mu_{jkl'} = 0$ vs

Ha: $\mu_{jkl} - \mu_{jk'l'} \neq 0$ or Ha: $\mu_{jkl} - \mu_{jk'l} \neq 0$ or Ha: $\mu_{jkl} - \mu_{jkl'} \neq 0$

for all j and $k \neq k'$ and $l \neq l'$

$$V(\bar{y}_{.jkl} - \bar{y}_{.jk'l'}) = V(\mu_{jkl} + \bar{\rho}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.jkl} - \mu_{jk'l'} - \bar{\rho}_{.} - \bar{\delta}_{.j} - \bar{\varepsilon}_{.jk'l'})$$

$$= V(\bar{\varepsilon}_{.jkl} - \bar{\varepsilon}_{.jk'l'})$$

$$= 2\sigma_e^2/r$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.jkl} - \bar{y}_{.jk'l'})} = 2MSE(bc)/r$$

ii) different levels of A with different (or same) levels of B*C

$Ho: \mu_{jkl} - \mu_{j'kl} = 0$ or $Ho: \mu_{jkl} - \mu_{j'k'l} = 0$ or

$Ho: \mu_{jkl} - \mu_{j'kl'} = 0$ or $Ho: \mu_{jkl} - \mu_{j'k'l'} = 0$ versus

$Ha: \mu_{jkl} - \mu_{j'kl} \neq 0$ or $Ha: \mu_{jkl} - \mu_{j'k'l} \neq 0$ or

$Ha: \mu_{jkl} - \mu_{j'kl'} \neq 0$ or $Ha: \mu_{jkl} - \mu_{j'k'l'} \neq 0$

for $j \neq j'$ and for all k, k', l and l'

$$V(\bar{y}_{.jkl} - \bar{y}_{.j'kl}) = V(\mu_{jkl} + \bar{\rho}_{.} + \bar{\delta}_{.j} + \bar{\varepsilon}_{.jkl} - \mu_{j'kl} - \bar{\rho}_{.} - \bar{\delta}_{.j'} - \bar{\varepsilon}_{.j'kl})$$

$$= V(\bar{\delta}_{.j} - \bar{\delta}_{.j'}) + V(\bar{\varepsilon}_{.jkl} - \bar{\varepsilon}_{.j'kl})$$

$$= 2\sigma_d^2/r + 2\sigma_e^2/r$$

$$= 2(\sigma_e^2 + \sigma_d^2)/r$$

The estimate of this variance is

$$\widehat{V(\bar{y}_{.jkl} - \bar{y}_{.j'kl})} = 2[MSE(a) + (bc-1)MSE(bc)]/bcr$$

Using the estimates of the variance above, the LSD value can be evaluated as follows:

a) between two A means

$$LSD_{\alpha} = t_{\alpha/2,(a-1)(r-1)} \sqrt{2MSE(a)/bcr}$$

b) betweeen two B means

$$LSD_{\alpha} = t_{\alpha/2,a(bc-1)(r-1)} \sqrt{2MSE(bc)/acr}$$

c) between two C means

$$LSD_{\alpha} = t_{\alpha/2,a(bc-1)(r-1)} \sqrt{2MSE(bc)/abr}$$

d) between two A*B means

i) with same level of A

$$LSD_{\alpha} = t_{\alpha/2,a(bc-1)(r-1)} \sqrt{2MSE(bc)/cr}$$

ii) with different levels of A

$$LSD_{\alpha} = t_1 \sqrt{2[MSE(a) + (b-1)MSE(bc)]/bcr}$$

e) between two A*C means

i) with same level of A

$$LSD_{\alpha} = t_{\alpha/2,a(bc-1)(r-1)} \sqrt{2MSE(bc)/br}$$

ii) with different levels of A

$$LSD_{\alpha} = t_2 \sqrt{2[MSE(a) + (c-1)MSE(bc)]/bcr}$$

f) between two B*C means

$$LSD_{\alpha} = t_{\alpha/2,a(bc-1)(r-1)} \sqrt{2MSE(bc)/ar}$$

g) between two A*B*C means

i) with same level of A

$$LSD_{\alpha} = t_{a/2,a(bc-1)(r-1)} \sqrt{2MSE(bc)/r}$$

ii) with different levels of A

$$LSD_{\alpha} = t_3 \sqrt{2[MSE(a) + (bc-1)MSE(bc)]/bcr}$$

where:
$$t_1 = \frac{t_{\alpha/2,(a-1)(r-1)}MSE(a) + t_{\alpha/2,a(bc-1)(r-1)}(b-1)MSE(bc)}{MSE(a) + (b-1)MSE(bc)}$$

$$t_2 = \frac{t_{\alpha/2,(a-1)(r-1)}\text{MSE(a)} + t_{\alpha/2,a(bc-1)(r-1)}(c-1)\text{MSE(bc)}}{\text{MSE(a)} + (c-1)\text{MSE(bc)}}$$

$$t_3 = \frac{t_{\alpha/2,(a-1)(r-1)}\text{MSE(a)} + t_{\alpha/2,a(bc-1)(r-1)}(ab-1)\text{MSE(bc)}}{\text{MSE(a)} + (ab-1)\text{MSE(bc)}}$$

IV. EFFECT OF INCORRECT ANALYSIS ON EXPERIMENT ARRANGED IN SPLIT PLOT.

(1) Using sub-plot error to test whole plot effects.

Consider an experiment involving r number of blocks, a levels of factor A (whole plot treatments) arranged in a randomized complete block design and b levels of factor B (sub-plot treatments) arranged at random with respect to the experimental units within each whole plot. The mathematical model for this experiment is defined by equation (1).

The expected mean squares (Table 1) enables us to determine the appropriate error term to be used as the denominator in testing the significance of the main effects and interaction effects. Under the null hypothesis of no factor A effects ($\alpha_j = 0$) the two expected mean squares

$$A \quad : \quad \sigma_e^2 + b\sigma_d^2 + Q(A) \qquad \text{and}$$

$$\text{Error(a)}: \quad \sigma_e^2 + b\sigma_d^2$$

are identical and hence a valid test is possible.

Suppose that we ignore the presence of the whole plot error term and instead we use the sub-plot error term as our denominator to test for the significance of factor A. Then, the two expected mean squares

$$A \quad : \sigma_e^2 + \sigma_d^2 + Q(A)$$

$$\text{Error(b)}: \sigma_e^2$$

will form the ratio $(\sigma_e^2 + b\sigma_d^2 + Q(A))/\sigma_e^2$. If there is no whole plot treatment effects ($\alpha_j = 0$) the ratio simply becomes $(\sigma_e^2 + b\sigma_d^2)/\sigma_e^2$ which is just the ratio of MSE(a) to MSE(b). Thus, the severity of the whole plot treatment effects basically depend on the magnitude of the whole plot error variance relative to the magnitude of the sub-plot error variance. A rough calculation will serve to illustrate how a relatively small value of σ_d^2 can grossly inflate Type I error rate of the whole plot effect. To this end, note that the ratio of mean squares which form the test statistic will have roughly 50 per cent chance of falling above the ratio of expected mean squares since the latter ratio will be near the center of the distribution of the test statistic.

Thus, an analyst who uses the sub-plot error term to test whole plot treatment effects will find critical values from the F-table with (a-1) and a(b-1)(r-1) degrees of freedom for the numerator and the denominator, respectively. If there are no whole plot treatment effect ($\alpha_j = 0$), the value of σ_d^2/σ_e^2 that causes the ratio of the expected mean squares to be equal to the 5 percent critical value in the F-table will be the value that causes the Type I error rate to be inflated from the stated 5% level to the 50% level, roughly. We find

$$\frac{\sigma_d^2}{\sigma_e^2} = \frac{F_{.05} - 1}{b} \qquad (6)$$

where $F_{.05}$ is the 5 % critical value from the F-table with (a-1) and a(b-1)(r-1) degrees of freedom for the numerator and denominator, respectively.

For simplicity, assume that there are 5 blocks, 3 levels of factor A and 4 levels of factor B, then

$$\frac{\sigma_d^2}{\sigma_e^2} = 0.565$$

This shows that if the component of variance due to whole plot is about 56% the sub-plot error variance, the Type I error rate will be inflated from 5% to around 50%. Similarly, if there are 3 levels of factor A, $r \geqslant 5$ and $b \geqslant 10$, then the value of σ_d^2 just 20% of σ_e^2 will cause a gross inflation of Type I error rate.

(2) Analyzing split-plot experiments using Randomized Complete Block analysis.

If we carry out the analysis of the experiment using the randomized complete block design instead of a split-plot design as specified in the given model, that is, ignoring the whole plot error term, then the following changes in the expected mean squares will follow:

Table 6. Analysis of Variance table of RCB when the model is Split-Plot.

Source	df	SS	MS	E(MS)
Replication	r-1	SSR	MSR	$\sigma_e^2 + b\sigma_d^2 + ab\sigma_r^2$
A	a-1	SSA	MSA	$\sigma_e^2 + b\sigma_d^2 + Q(A)$
B	b-1	SSB	MSB	$\sigma_e^2 + Q(B)$
A*B	(a-1)(b-1)	SSAB	MSAB	$\sigma_e^2 + Q(AB)$
Error	(ab-1)(r-1)	SSE	MSE	$\sigma_e^2 + b(a-1)\sigma_d^2/(ab-1)$
Total	abr-1	SST		

From the above table and forming the ratio to test for the significance of the main effects and interaction effects, we use the error mean square as the denominator. The critical values from the F-table will be (a-1) and (ab-1)(r-1), (b-1) and (ab-1)(r-1), and (a-1)(b-1) and (ab-1)(r-1) degrees of freedom for the numerator and denominator for factor A effects, factor B effects and interaction effects, respectively. Likewise, looking closely on the effect of factor A, assuming that there is no factor A effect ($\alpha_j = 0$), then the ratio of factor A to the error term or correspondingly the whole plot comparison would inflate the Type I error rate. That is, the ratio of E(MSA)/E(MSE), under the null hypothesis of no A effect, would be

$$\frac{\sigma_e^2 + b\sigma_d^2}{\sigma_e^2 + b(a-1)\sigma_d^2/(ab-1)} \tag{7}$$

Note that $b > \dfrac{b(a-1)}{(ab-1)}$ so that the expected mean squares is greater than 1. Setting this value equal to the 5% critical value and solving for σ_d^2/σ_e^2 we find

$$\frac{\sigma_d^2}{\sigma_e^2} = \frac{F_{.05} - 1}{b - F_{.05}(a-1)b/(ab-1)} \tag{8}$$

where $F_{.05}$ is the 5% critical value from the F-table with (a-1) and (ab-1)(r-1) degrees of freedom for numerator and denominator, respectively.

Suppose we have 5 replications, 3 levels of factor A and 5 levels of factor B, then ,

$$\frac{\sigma_d^2}{\sigma_e^2} = \frac{F_{.05,2,64} - 1}{4 - (10/14)F_{.05,2,64}} = 0.79$$

This shows that if the component of variance due to whole plot is about 79% the error variance, then the Type I error rate will be inflated from 5% to around 50%.

Similarly, if there are 3 levels of factor A, $r \geqslant 5$ and $b \geqslant 10$, then the value of σ_d^2 just 22% of σ_e^2 will cause a gross inflation of Type I error rate.

For comparisons of Factor B and the interaction effects, however, there is a decrease in efficiency using this type of analysis when the experimental design is a split-plot. This is to be expected since the error variance for an RCB analysis is increased by an amount $b(a-1)\sigma_d^2/ab-1$ hence causing too few rejections of the null hypotheses of no B effects and interaction effects.

(3) <u>Computer Simulation.</u>

Using the Scientific Subroutine Package (SSP) and Statistical Analysis System (SAS), a computer simulation was conducted where the defined model is a split-plot design model. Five replications, 3 levels of factor A and 4 levels of factor B were considered in this study. The main effects and the interaction effects were assumed to be insignificant in generating the observations. The data was generated using SSP and the analysis was done using SAS. The procedure was run 40 times.

a) Using sub-plot error to test the whole plot effects.

Using the data generated, analysis was done using the split plot model but uses the sub-plot error term to test for the significance of factor A. Results shows that a ratio of σ_d^2/σ_e^2 of about 0.56 inflates the Type I error rate from 5% to 47.5%. This indicate that using the sub-plot error term to test the whole plot effects would lead to an incorrect results.

b) Using RCB analysis.

Using the same data generated from a split plot design experiment, a Randomized Complete Block analysis was done. Results shows that using the 'weighted' error variance to test for the whole plot effects inflates the Type I error rate from 5% to 35%.

In general, analyzing experiment using the inappropriate error term from experiment which are of a split plot arrangement will cause factor A to be significant at most 50% of the time and thus would lead to an erroneous conclusion about the real effect of factor A.

V. REPEATED MEASURES DESIGN

(1) Repeated Measures Design Defined.

An experimental design in which experimental units are used repeatedly by exposing them to a sequence of different or identical treatments is called a repeated measurement design. These types of design are extensively used in agricultural, industrial and psychological research.

Several types of experiment which are of a repeated measures design are as follows.

a) Suppose we have two levels of factor A and two levels of factor B and we are interested on the effects of A, B, and A*B on an individual (experimental units). We select, say 4 individuals, and then give all the four treatment combinations to an individual in a random order one at a time. The experimental lay-out, after randomization, might look as

Subject	1	A_1B_1	A_2B_1	A_1B_2	A_2B_2
	2	A_2B_1	A_1B_2	A_1B_1	A_2B_2
	3	A_2B_2	A_1B_1	A_2B_1	A_1B_2
	4	A_1B_2	A_2B_2	A_2B_1	A_1B_1

Treating the subjects as blocks, then we might analyze this experiment just like a Randomized Complete Block.

b) The second type of experiment can be illustrated as follows. Suppose we would like to evaluate the effect of two treatments, say treatment A and treatment B, and also we would like to include possible order effects. If there are 10 subjects, then we can assign treatment A followed by treatment B to 5 subjects and the remaining subjects receiving treatment B first followed by treatment A. The lay-out might look like as follows:

		Subjects									
		1	2	3	4	5	6	7	8	9	10
	1	A	B	A	B	B	A	B	A	B	A
Order											
	2	B	A	B	A	A	B	A	B	A	B

This experiment is an example of a two-period Cross-over design.

c) The third type of repeated measures is of a split-plot type. Suppose we have a levels of factor A and b levels of factor B. The experimental subjects are assigned randomly to one and only one level of factor A. After the subject is assigned to a particular level of factor A, then all levels of factor B will be assigned one after another in random order to that subject.

Considering this type of arrangement, we might analyze this experiment using a split plot analysis where subjects and time periods are the main plot and sub-plots and factor A and factor B as the main plot treatments and sub-plot treatments, respectively.

d) The fourth type are experiments in which measurements are obtained over time. Suppose we have t treatments and r experimental units. By measuring the experimental units at several times, the experimental unit is essentially being 'split' into parts (time interval) and the response is being measured on each part. The larger experimental unit is the subject or the collection of time intervals while the smaller experimental unit is the interval of time during which the subject is exposed to a treatment or an interval just between time measurements. The treatment is the main plot treatments and the time is the sub-plot treatments.

So far, we have defined some experiments where we can make use of the repeated measures design. The need for these designs can be justified in several ways.

a) Due to budget limitation, the experimenter has to use each experimental unit for several tests.

b) In some experiments the treatment effects do not have a serious damaging effect on the experimental unit and therefore these experimental units can be used for successive experiments.

c) In some experiments, the experimental units are human beings or animals and often the nature of the experiment is such that it calls for special training over a long period of time. Therefore, due to time limitation, one is forced to use these experimental units for several tests.

d) One of the objectives of the experiment is to find out the effects of the different sequences as in drug, nutrition, or learning experiments.

e) Sometimes the experimental units are scarce, therefore the experimental units have to be used repeatedly.

(2) Split Plot Type Repeated Measures Design

In comparison to the usual split plot, repeated measurement design may not allow randomization of one of the experimental units. If one of the factors is time, then the levels of time cannot be randomly assigned to the time intervals and thus the usual analysis of variance for split plot may not be valid. Because of this non random assignment, the errors corresponding to the experimental unit may have a covariance matrix that does not conform to those for which the usual split plot analysis is valid.

Suppose there are p levels of some treatment, with r subjects exposed to the ith level and each subject measured on some characteristics (say time) periodically for t times post treatment. This experiment can be described as having two experimental units where the larger size of experimental unit is the subject and the smaller experimental unit is the time interval.

A model to describe the response that reflects the two sizes of experimental unit is

$$y_{ijk} = \mu + \alpha_i + \delta_{j(i)} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk} \tag{9}$$

$$i=1,2,\ldots,p \quad j=1,2,\ldots,r \quad k=1,2,\ldots,t$$

where μ = overall mean

α_i = effect of the ith treatment level

$\delta_{j(i)}$ = effect of the jth individual within the ith treatment level

β_k = kth time effect

$(\alpha\beta)_{ik}$ = treatment by time interaction effect

ε_{ijk} = time interval errors within subject

The assumptions often made for an analysis of the above model are

(1) $\delta_{ij} \sim N(0,\sigma_d^2)$

(2) $\varepsilon_{ijk} \sim N(0,\sigma_e^2)$

(3) δ_{ij} and ε_{ijk} are independent

(4) Treatments and time factors are fixed

(5) $\sum_i \alpha_i = \sum_k \beta_k = \sum_i (\alpha\beta)_{ik} = \sum_k (\alpha\beta)_{ik} = 0$

From the above equation, $\mu + \alpha_i + \delta_{ij}$ represent the subject part of the model and $\beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}$ represent the time interval part of the model.

In any statistical analysis, account must be taken of the fact that the observation from different sub-units in the same unit may be correlated. To carry such an analysis, assume that a correlation, ρ, exists between the experimental error for any 2 sub-units in the same unit and that the experimental error for any two sub-units in different units are assumed to be uncorrelated. Hence, define the covariance $(k \neq k')$

$$E(\varepsilon_{ijk}\varepsilon_{ijk'}) = \text{constant} \qquad \text{for all pairs } k,\ k'$$

and the variance for each sub-plot treatment

$$E(\varepsilon_{ijk}^2) = E(\varepsilon_{ijk'}^2) = \sigma^2 \tag{10}$$

then the correlation can be written as

$$\rho = \frac{E(\varepsilon_{ijk}\varepsilon_{ijk'})}{\sqrt{E(\varepsilon_{ijk}^2)\; E(\varepsilon_{ijk'}^2)}} = \frac{E(\varepsilon_{ijk}\varepsilon_{ijk'})}{\sigma^2} \qquad (11)$$

thus, $$E(\varepsilon_{ijk}\varepsilon_{ijk'}) = \rho\sigma^2 \qquad (12)$$

Using this relationship, then the covariance matrix for observations within a subject is thus assumed to have the form

$$\sum = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \vdots & & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} \qquad (13)$$

Likewise, for a random variable y_{ijk} which is assumed to have a normal distribution with mean zero, the variance from model (9) is found to be $\sigma_d^2 + \sigma_e^2$ and the covariance is

$$\mathrm{Cov}\,(y_{ijk}y_{ijk'}) \begin{cases} = \sigma_d^2 + \sigma_e^2 & i = i' \quad j = j' \quad k = k' \\ = \sigma_d^2 + \rho\sigma^2 & i = i' \quad j = j' \quad k \neq k' \\ = 0 & \text{otherwise} \end{cases} \qquad (14)$$

It can be shown (Curnow,1957, Greenhouse and Geisser, 1959, Danford, Hughes and McNee, 1960) under this kind of covariance matrix that the usual F-tests for split-plot design is valid.

However, before doing the analysis part, we must be sure that the validity of the assumption that the variances and covariances are the same over the various treatments are checked for each experimental condition. The procedure for testing this assumption has been given by Box (1950) as an extension to the multivariate situation of Bartlett's homogeneity of variance test (see Box, 1949). To test the hypothesis that the covariance matrices S_1, S_2, ..., S_p are random samples for populations in which the covariance matrices are equal, that is, $\sum_1 = \sum_2 = \dots = \sum_p = \sum$, one uses the statistic

$$M_1 = N \ln |S| - \Sigma r \ln |S_i| \qquad (15)$$

$$C = \frac{2t^2 + 3t - 1}{6(t+1)(p-1)} \Sigma [(1/r) - 1/N] \qquad (16)$$

$$f_1 = t(t+1)(p-1)/2 \qquad (17)$$

where $N = \Sigma r$, the total number of subjects

S = pooled covariance matrix

S_i = matrix of covariances for level a_i

t = number of time inteval

p = number of levels of factor A

Under the hypothesis that the multivariate normal populations have equal covariances then the statistic

$$X_1^2 = (1 - C_1)M_1 \tag{18}$$

has a sampling distribution which is approximated by a chi-square distribution having f_1 degrees of freedom.

A matrix having compound symmetry has the form of Eqn.(13). The compound symmetry condition implies that the random variables are equally correlated and have equal variances. Huynh and Feldt (1970) gave a more general form of the covariance matrix. The Huynh and Feldt condition specifies that the elements of a covariance matrix $\sum$ be expressed as

$$\sigma_{tt} = \gamma_t + \gamma_{t'} + \lambda\delta_{tt'} \tag{19}$$

$$\text{where, } \delta_{tt'} = 1 \quad \text{if } t = t' \tag{20}$$

$$= 0 \quad \text{if } t \neq t'$$

or in terms of matrices

$$\sum = \lambda I + \gamma \underline{j}' + \underline{j}\gamma' \tag{21}$$

where I is a t x t identity matrix

$\underline{j}$ is a t x 1 vector of ones

$\gamma' = (\gamma_1, \gamma_2, \ldots, \gamma_t)$ are unknown parameters

The Huynh and Feldt condition is equivalent to specifying that the variances of the difference between pairs of errors, such as $\varepsilon_{ijk} - \varepsilon_{ijk'}$, are equal for all k and k', k ≠ k'. If the variances are all equal, then the condition is equivalent to compound symmetry.

To test the hypothesis that the covariance matrix, $\sum$, satisfies the Huynh and Feldt condition, one uses the following statistic.

$$M_2 = -v \ln(\hat{\theta}) \tag{22}$$

$$C_2 = \frac{t(t+1)^2(2t-3)}{6(v)(t-1)(t^2+t-4)} \tag{23}$$

$$f_2 = (t^2 + t - 4)/2 \tag{24}$$

where $v = a(r-1)(t-1)$

$$\theta = \frac{t^2(\bar{\sigma}_{ii} - \bar{\sigma}_{..})^2}{(t-1)[\,\Sigma\Sigma\, \sigma_{ij}^2 - 2t\Sigma\, \bar{\sigma}_{i.}^2 + (t\bar{\sigma}_{..})^2]} \tag{25}$$

where σ_{ij} are the elements of $\sum$

$\bar{\sigma}_{..} = \Sigma\Sigma\, \sigma_{ij}/t^2$

$\bar{\sigma}_{i.} = \Sigma\, \sigma_{ij}/t$

$\bar{\sigma}_{ii} = \Sigma\, \sigma_{ii}/t$

An estimate of $\sum$ can be obtained using the relation

$$\pi = [I_t - (1/t)J_t] \hat{\sum} [I_t - (1/t)J_t] \tag{26}$$

where I_t = is a t x t identity matrix

J_t = is a t x t matrix of ones

Because of the form of π, the value of θ is simplified as

$$\theta_\pi = \frac{t^2(\bar{\pi}_{ii})^2}{(t-1) \sum\sum \pi_{ij}^2} \tag{27}$$

where $\bar{\pi}_{ii}$ is the mean of the diagonal elements of π and

π_{ij} is the ijth element of the matrix π

Under the null hypothesis that $\sum$ has the specified form, the statistic

$$X_2^2 = (1 - C_2)M_2 \tag{28}$$

has a sampling distribution which can be approximated by a chi-square distribution having f_2 degrees of freedom.

Thus, we say that $\sum$ has a compound symmetry if $X_2^2 < X^2_{a,f_2\ df}$ and we proceed to analyze the data just like a split plot design. However, if $X_2^2 > X^2_{a,f_2\ df}$ then the usual analysis for split plot design is not valid. Hence an appropriate solution is the one suggested by Box (1954). That is, use θ as a correction factor for adjusting the degrees of freedom to those sources of variation that are based on the within-subject comparisons. No adjustment is necessary for the between-subject comparisons. Box shows that for all $t \geqslant 2$ the range of θ is $1/(t-1) \leqslant \theta \leqslant 1$ with $\theta = 1$ when $\sum$ satisfies the Huyn and Feldt condition.

Computing the value of θ in order to adjust the degrees of freedom is not an easy task. Greenhouse and Geisser (1959) developed a three-step procedure that can prevent having to compute the value of θ. The procedure are as follows.

a) Compare the F-ratio in question to the percentage point with the usual degrees of freedom. If it is not significant, stop. The adjusted degrees of freedom test will also be not significant.

b) Compare the F-ratio in question to the percentage point with the Conservative Box correction degrees of freedom. If it is significant, stop. The adjusted degrees of freedom test will also be significant. (where the Conservative Box correction degrees of freedom is to divide each of the respective usual degrees of freedom by (t-1).

c) If the F-ratio is significant with the usual degrees of freedom and not significant with the conservative Box correction degrees of freedom, then the θ adjusted degrees of freedom must be estimated to make a decision.

Table 7 shows the analysis of variance table for repeated measures design when the covariance matrix has the form of compound symmetry.

Table 7. Analysis of Variance Table for Repeated Measures Design (Split-plot type analysis)

Source	df	SS	MS	E(MS)
A	a-1	SSA	MSA	$\sigma_e^2(1 + 2\lambda) + t\sigma_d^2 + Q(A)$
Error(a)	r(a-1)	SSE(a)	MSE(a)	$\sigma_e^2(1 + 2\lambda) + t\sigma_d^2$
Time	t-1	SST	MST	$\sigma_e^2 + Q(T)$
A*Time	(a-1)(t-1)	SSA*T	MSA*T	$\sigma_e^2 + Q(A{*}T)$
Error(b)	a(r-1)(t-1)	SSE(b)	MSE(b)	σ_e^2
Total	atr - 1	SSTotal		

The term λ occurs because the variance of $\bar{\varepsilon}_{ij.}$, the mean of the errors of subject j assigned to the i<u>th</u> level of factor A, is

$$\mathrm{Var}(\bar{\varepsilon}_{ij.}) = \sigma_e^2(1 + 2\lambda)/t \tag{29}$$

(3) <u>Autocorrelated errors.</u>

Now, suppose that the covariance matrix does not conform to a compound symmetry but instead the errors are correlated through an autocorrelated structure, then the assumption corresponding to the within-subject error structure will be changed. Further, suppose that the error structure follows a first-order autoregressive model, that is,

$$\varepsilon_{ijk} = \rho\varepsilon_{ijk-1} + z_{ijk} \tag{30}$$

then the model for a split-plot repeated measure design will just be the same model as before with an added assumption that

$$\varepsilon_{ij0} \sim N[0,\sigma_z^2/(1-\rho^2)] \quad \text{and}$$

$$z_{ijk} \sim N(0,\sigma_z^2)$$

The corresponding covariance matrix for the errors associated with the within subject data obtained in a repeated measures design having a first-order autocorrelated error will be,

$$\sum = \frac{\sigma^2}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{t-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{t-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{t-3} \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \rho^{t-1} & \rho^{t-2} & \rho^{t-3} & \rho^{t-4} & \cdots & 1 \end{bmatrix} \tag{31}$$

This topic will be limited at this point and will not be discussed further. For different types of time series models which might be appropriate for describing the within-subject error structure, readers are advised to consult the book of Box and Jenkins on Time Series Analysis.

In particular see Albohali's (1) result for the first order autocorrelation error structure.

(4) Numerical Example.

Suppose we are given the following data:

			Time	
Treat	Subject	T1	T2	T3
	1	4	7	3
	2	3	5	1
1	3	7	9	6
	4	6	6	2
	5	5	5	1
	6	8	2	5
	7	4	1	1
2	8	6	3	4
	9	9	5	2
	10	7	1	1

First we test the hypothesis,

$$Ho: \Sigma_1 = \Sigma_2 = \dots = \Sigma_p = \Sigma$$

$$Ha: \Sigma_1 \neq \Sigma_2 \neq \dots \neq \Sigma_p \neq \Sigma$$

We compute the sample covariance matrix for each level of factor A and the pooled covariance matrix, that is,

$$a_1$$

$$S_1 = \begin{array}{cccc} & T1 & T2 & T3 \\ T1 & 2.50 & 1.75 & 2.50 \\ T2 & & 2.80 & 3.30 \\ T3 & & & 4.30 \end{array} \qquad |S_1| = 1.08$$

$$a_2$$

$$S_2 = \begin{array}{c|ccc} & T1 & T2 & T3 \\ \hline T1 & 3.70 & 2.10 & 1.15 \\ T2 & & 2.80 & 0.70 \\ T3 & & & 3.30 \end{array} \qquad |S_2| = 17.51$$

$$S = \begin{array}{c|ccc} & T1 & T2 & T3 \\ \hline T1 & 3.10 & 1.92 & 1.82 \\ T2 & & 2.80 & 2.00 \\ T3 & & & 3.80 \end{array} \qquad |S| = 11.28$$

The statistics defined in equation 15 to equation 18 are obtained as follows

$$M_1 = 10\ln(11.28) - 5\ln(1.08) - 5\ln(17.51) = 9.53$$

$$C_1 = \frac{2(3)^2+3(3)-1}{6(3+1)(2-1)} \quad [(1/5 + 1/5) - (1/10)] = 0.325$$

$$f_1 = 3(3+1)(2-1)/2 = 6$$

$$X^2 = (1-0.325)9.53 = 6.43$$

$$X^2_{.05,6} = 12.6$$

Since $6.43 < 12.6$ then the hypothesis of homogeneity of covariances are accepted.

Next, we test the hypothesis that $\sum$ has the form of compound symmetry. Using the estimate of $\sum$ that is π , we obtain the following quantities:

$$
\pi = \begin{array}{cccc} & T1 & T2 & T3 \\ T1 & 0.8933 & -0.2467 & -0.6465 \\ T2 & & 0.6733 & -0.4267 \\ T3 & & & 1.0733 \end{array}
$$

From this estimate of the covariance matrix, we obtain

$$\theta_{\pi} = 9(0.88)^2/2(3.7253) = 0.9354$$

$$M_2 = -16\ln(.9354) = 1.0685$$

$$C_2 = \frac{3(3+1)^2(2*3-3)}{6(16)(3-1)(3^2+3-4)} = 0.0938$$

$$f_2 = (3^2+3-4)/2 = 4$$

$$X^2 = 0.97$$

$$X^2_{.05,4} = 9.5$$

Since 0.97 < 9.5, then we accept the hypothesis that the covariance matrix follows the form of a compound symmetry.

Now, since we cannot reject the hypothesis that the covariance matrix has the form of a compound symmetry, then we analyze the set of data using the usual split-plot analysis. The following table follows:

Table 8. Analysis of Variance Table of the Data Set.

Source	df	Sum of Squares	Mean Squares	F-value
A	1	3.33	3.33	1.95
Error(a)	8	13.67	1.71	
Time	2	58.07	29.04	7.26*
A*Time	2	44.87	22.44	5.61*
Error(b)	16	63.94	4.00	

From the table note that there is a significant A*Time interaction, thus, we need to compare the times at each level of factor A and the levels of factor A at each time.

To compare times at each level of factor A, the standard error of the difference of the two means is

$$\begin{aligned} s.e.(\bar{y}_{i.k} - \bar{y}_{i.k'}) &= \sqrt{2MSE(b)/r} \\ &= \sqrt{2(4.00)/3} \\ &= 1.63 \end{aligned}$$

and a 5% LSD value is

$$\begin{aligned} LSD_{.05} &= t_{.025,16}(s.e) \\ &= 2.120(1.63) \\ &= 3.46 \end{aligned}$$

A comparison of the time means within a level of Factor is shown in Table 9.

Table 9. Comparisons of Time means at same level of factor A.

Time	Factor A a_1	Factor A a_2
T1	5.00(ab)	6.80(a)
T2	6.40(a)	2.40(b)
T3	2.40(b)	2.60(b)

Note: Means within a column with the same letter are not significantly different at 5% level.

Suppose that $\sum$ does not follow the compound symmetry form, then θ adjusted degrees of freedom should be applied to test for the significance of the sub-plot treatment effects and the interaction effcts. Box showed that F_T is approximately distributed as an F-distribution with $\theta(t-1)$ and $\theta(a)(t-1)(r-1)$ degrees of freedom for the numerator and denominator, respectively. Table 10 shows the actual and θ adjusted degrees of freedom for the given data, assuming that $\sum$ is not of a form of compound symmetry.

Table 10. Actual and Adjusted Degrees of Freedom.

Source	Actual	θ adjusted df (θ = 0.9357)
A	1	1*
Error(a)	8	8*
Time	2	1.87
A*Time	2	1.87
Error(b)	16	14.97

*Unchanged (between-person comparison)

Notice that using the θ adjusted degrees of freedom results in the same conclusion as before.

VI. SUMMARY

Occasionally some factors in an experiment can be applied differentially to smaller units than can others. Dietary comparisons must be made on whole animals, whereas drugs can sometimes be compared by injection at different sites on one animal. The comparison of soil-cultivation techniques that employ unwieldy implements may demand large plots, but tests of fertilizers or other agronomic factors may be made simultaneously on subdivision of these areas. An experiment in which some treatments are applied to large units, or whole plots, each of which are divided into two or more sub-plots for other treatments, is said to have a split-plot design.

Split-plot experiments will usually assess the effects of sub-factors and their interaction with whole plot factors more precisely than the effects of whole plot factors alone. Thus, split-plot designs are adopted in order to obtain higher precision on comparisons of greater importance.

The split-plot design lends itself readily to various modifications. Two whole plot factors with single sub-plot factor and single whole plot factor with two sub-plot factors were considered in this study.

The expected values of sample mean squares in terms of population parameters were derived such that appropriate error terms are used for testing various hypotheses. Likewise, standard errors were determined to be used in different treatment comparisons

A simulation study was also conducted to find out the effect of an incorrect analysis on experiment arranged in split-plot. Results show that even a small ratio of σ_d^2/σ_e^2 that causes the expected mean squares to be equal to the 5% in the F-table will be the value that causes the Type I error rate to be inflated from the stated 5% to the 50% level roughly when the incorrect error term is used to test the effect of the whole plot factor.

Experiments where the experimental units are used repeatedly can also be analyzed using the usual split-plot analysis provided that the covariance matrix for observations within a subjct has a compound symmetry form. However, if the covariance matrix do not conform to compound symmetry, then the Θ adjusted degrees of freedom should be applied to test for the significance of the sub-plot treatment effects and the interaction effects.

REFERENCES

1. ALBOHALI, M. (1983). A Time Series Approach to the Analysis of Repeated Measures Design. Ph.D. Dissertation, KSU, Manhattan, Kansas.

2. ANDERSON, R.L. and T.A. BANCROFT (1952). Statistical Theory in Research. New York. McGraw-Hill.

3. BENNETT, C.A. and N.L. FRANKLIN (1954). Statistical Analysis in Chemistry and the Chemical Industry. Wiley and Sons, Inc., New York.

4. BOX, G.E.P. (1949). A General Dsitribution Theory for a Class of Likelihood Criteria. Biometrika, 36 : 317-46.

5. __________ (1950). Problems in the Analysis of Growth and Wear Curves. Biometrics, 6 : 362-89.

6. __________ (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems. Annals of Mathematical Statistics, 25 : 290-302.

7. BLAIR, R.C., J.J. HIGGINS, M.E.H. TOPPING and A.L. MORTIMER (1983). An Investigation of the Robustness of the _t_ Test to Unit of Analysis Violations. Educational and Psychological Measurement.

8. COCHRAN, W.G. and G.M. COX (1957). Experimental Designs. 2nd ed. John Wiley and Sons, Inc., New York.

9. COLLIER, R.O.,Jr., F.B. BAKER, G.K. MANDEVILLE and T.F. HAYES (1967). Estimates of Test Size for Several Test Procedures based on Conventional Variance Ratios in the Repeated Measures Design. Psychometrika, 32 : 339-53.

10. CORNFIELD, J. and J.W. TUKEY (1956). Average Values of Mean Squares in Factorials. Annals of Mathematical Statistics, 27 : 907-49.

11. CURNOW, R.N. (1957). Heterogeneous Error Variances in Split-Plot Experiments. Biometrika, 44: 378-83.

12. DANFORD, M.B., H.M. HUGHES and R.C. McNEE (1960). On the Analysis of Repeated Measurement Experiments. Biometrics, 42 : 547-65.

13. FRYER, H.C. (1966). Concepts and Methods of Experimental Statistics. Allyn and Bacon, Inc., Boston.

14. GREENHOUSE, S.W. and S. GEISSER (1959). On Methods in the Analysis of Profile Data. Psychometrika, 24 : 95-112.

15. HEDAYAT, A. and K. AFSARINEJAD (1975). A Survey of Statistical Design and Linear Models . _Repeated Measurements Designs, I_. North - Holland Publishing Co., 229-42.

16. HUYNH, H. and L.S. FELDT (1970). Conditions Under Which Mean Square Ratios in Repeated Measurement Designs Have Exact F-Distributions. Journal of the American Statistical Association, 65 : 1582-89.

17. LANA, E.R. and A. LUBIN (1963). The Effect of Correlation on the Repeated Measures Design. Educational and Psychological Measurement, XXIII : 729-39.

18. MILLIKEN, G.A and D.E. JOHNSON (1984). Analysis of Messy Data. Volume 1. Designed Experiments. Wadsworth, Inc.

19. NAMBOODIRI, N.K. (1972). Experimental Designs in Which Each Subject is Used Repeatedly. Psychological Bulletin, 77 : 54-64.

20. SAS Institute, Inc. (1982). SAS User's Guide: Statistics, 1982 edition. Cary, N.C., SAS Institute, Inc.

21. SATTERTHWAITE, F.E. (1946). An Approximate Distribution of Estimates of Variance Components. Biometrics Bulletin, 2 : 110-14.

22. WINER, B.J. (1962). Statistical Principles in Experimental Design. 2nd ed., McGraw-Hill Book Co, New York.

23. YATES, F. (1970). Experimental Design. Selected Papers. Charles Griffin and Co., Ltd.

APPENDIX

Program to generate observations from a Normal Distribution

```
      DIMENSION A(3),B(4),AB(3,4)
      IX=585226667
      IY=610986963
      S=SQRT(0.565)
      V=1.0
      AM=0.0
      DO 5 I=1,5
      DO 5 J=1,3
      A(J)=8
      CALL GAUSS (IX,S,AM,X1)
      DO 5 K=1,4
      B(K)=8
      AB(J,K)=8
      CALL GAUSS (IY,V,AM,Z1)
      Y1=A(J)+X1+B(K)+AB(J,K)+Z1
5     WRITE (15,500) I,J,K,Y1
500   FORMAT (1X,3I2,1X,F7.2)
      STOP
      END
```

This subroutine computes a normally distributed random numbers with a given mean and standard deviation.

```
      SUBROUTINE GAUSS (IX,S,AM,V)
      A=0.0
      DO 50 I=1,12
      CALL RANDU (IX,IY,Y)
      IX=IY
50    A=A+Y
      V=(1-6.0)*S+AM
      RETURN
      END
```

SAS Program for Split Plot Design

```
DATA ONE;
INPUT REP WP SP OBS;
CARDS;
        .
        .
   data cards
        .
        .
PROC ANOVA;
TITLE USUAL SPLIT-PLOT ANALYSIS OF VARIANCE;
CLASSES REP WP SP;
MODEL OBS=REP WP REP*WP SP WP*SP;
* RESULT OF THIS PROCEDURE GIVES US THE TEST OF THE
  WHOLE PLOT EFFECT USING THE SUB-PLOT ERROR TERM*;
TEST H=WP E=REP*WP;
* USING THE OPTION TEST H=WP E=REP*WP THE WHOLE PLOT EFFECT IS TESTED
  BY THE APPROPRIATE ERROR TERM (WHOLE PLOT ERROR)*;
PROC ANOVA,
TITLE USING RCB ANALYSIS FOR EXPERIMENT LAID OUT IN SPLIT-PLOT DESIGN;
CLASSES REP WP SP;
MODEL OBS=REP WP SP WP*SP;
```

SPLIT PLOT DESIGNS

by

MARCIAL PEREZ ESTOLANO

B.S.A., University of the Philippines at Los Baños, 1976

AN ABSTRACT OF MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1985

ABSTRACT

Split-plot designs are used in two-factor experiments where one factor requires larger experimental units than the other. These kind of experiments will assess the effects of sub-factors and its interactions more precisely than the effects of whole plot factors.

Expectation of mean squares were derived for the simple split-plot design and two of its variants. Likewise, standard errors for multiple comparisons were determined.

Two types of incorrect analysis, namely, (1) sub-plot error used for whole plot tests and (2) Randomized Complete Block analysis instead of Split Plot analysis were considered in this study to find out the effects on Type I error when the experiment is arranged in Split Plot Design. Using both ratios of expected mean squares and simulation, results showed that ignoring the whole plot error, even if it is small, can cause substantial inflation of Type I error rate.

Split-plot type repeated measures design can be analyzed using the usual split-plot analysis provided that the covariance matrix for observations within a subject has a compound symmetry form. However, an adjusted degrees of freedom should be applied when this covariance matrix does not have a compound symmetry. An example of a repeated measures analysis is given in which a preliminary test of compound symmetry is made.