

INCOMPLETE DATA IN SPLIT-PLOT, SPLIT-BLOCK, AND
OTHER SIMILAR DESIGNS*

BU-178-M

W. T. Federer

February, 1965

Abstract

The general form of analyses for split-plot, split-block, and similar designs with missing observations is presented through the medium of a specific example. In the example one of the treatments was deleted from a split-block design wherein one set of treatments was arranged in a 4 x 4 latin square and the other set of treatments in a randomized complete blocks design with four replicates; the columns of the 4 x 4 latin square design represented the complete blocks for the randomized complete blocks design. An alternate analysis utilizing dummy variates and a covariance analysis is discussed.

* Biometrics Unit, Cornell University, Ithaca, N. Y.

INCOMPLETE DATA IN SPLIT-PLOT, SPLIT-BLOCK, AND
OTHER SIMILAR DESIGNS*

BU-178-M

W. T. Federer

February, 1965

In the course of consulting the following design came under consideration:**

	b	a	d	c	a	c	d	b	d	b	a	c	a	b	c	d
Row 1		1				2				3				4		
Row 2		2				3				4				1		
Row 3		3				4				1				2		
Row 4		4				1				2				3		
	Column 1		Column 2		Column 3		Column 4									

Treatments 1, 2, 3, and 4 represent time of harvest and were laid out in a 4x4 latin square design. Treatments a, b, c, and d represent four varieties which were laid out in a randomized complete block design with the four columns of the latin square representing the four complete blocks or replicates for varieties.

* Biometrics Unit, Cornell University, Ithaca, N. Y.

** From L. A. Montoya, Vegetable Crops Department, Cornell University.

We note that the lay-out is a split block design with the four harvests in a latin square design and the four varieties in a randomized complete block design. Therefore, the yield equation is of the form:

$$Y_{ghij} = n_{ghij} (\mu + \rho_g + \lambda_h + \alpha_i + \epsilon_{ghi} + \beta_j + \alpha\beta_{ij} + \delta_{hj} + \pi_{ghij}),$$

where μ = an effect common to all observations, ρ_g = row effect, λ_h = column and complete block effect, α_i = harvest effect, β_j = variety effect, $\alpha\beta_{ij}$ = harvest by variety interaction effect, ϵ_{ghi} = randomly distributed error effects for harvests with mean zero and common variance σ_ϵ^2 , δ_{hj} = randomly distributed error effects for varieties with mean zero and common variance σ_δ^2 , π_{ghij} = randomly distributed error effects for harvest by variety interaction with mean zero and common variance σ_π^2 , $g = 1, 2, 3, 4 = k$, $h = 1, 2, 3, 4 = b$, $i = 1, 2, 3, 4 = v$, $j = 1, 2, 3, 4 = t$, and n_{ghij} = one if the ij^{th} harvest and variety combination occurs in the g^{th} row of the h^{th} column and = zero otherwise.

Let $Y_{\dots} = \sum_g \sum_h \sum_i \sum_j n_{ghij} Y_{ghij}$ = grand total, $Y_{g\dots}$ = g^{th} row total, $Y_{\dots h}$ = h^{th} column total, $Y_{\dots i}$ = i^{th} harvest total, $Y_{\dots j}$ = j^{th} variety total, $Y_{ghi} = \sum_j n_{ghij} Y_{ghij}$, $Y_{\dots ij}$ = total of ij^{th} harvest and variety combination, and $Y_{\dots h \cdot j}$ = total for variety j in column (complete block) h . Then, the key-out of degrees of freedom and the sums of squares in the analysis of variance with $bkt = 64$ observations are:

<u>Source of variation</u>	<u>d.f.</u>	<u>Sum of Squares</u>
Correction for mean	1	$Y_{\dots}^2 / 64$
Row	3	$\sum_{g=1}^4 \frac{Y_{g\dots}^2}{16} - \frac{Y_{\dots}^2}{64}$
Column	3	$\sum_{h=1}^4 \frac{Y_{\dots h}^2}{16} - \frac{Y_{\dots}^2}{64}$

Harvest	3	$\sum_{i=1}^4 \frac{Y_{\dots i}^2}{16} - \frac{Y_{\dots}^2}{64}$
Error for harvest	6	$\sum_g \sum_h \frac{Y_{gh\dots}^2}{4} - \frac{\sum Y_g^2 \dots}{16} - \frac{\sum Y_h^2 \dots}{16} - \frac{\sum Y_{\dots i}^2}{16} + \frac{2Y_{\dots}^2}{64}$
Variety	3	$\sum_{j=1}^4 \frac{Y_{\dots j}^2}{16} - \frac{Y_{\dots}^2}{64}$
Error for variety	9	$\sum_h \sum_j \frac{Y_{\dots h j}^2}{4} - \frac{\sum Y_h^2 \dots}{16} - \frac{\sum Y_{\dots j}^2}{16} + \frac{Y_{\dots}^2}{64}$
Variety x harvest	9	$\sum_i \sum_j \frac{Y_{\dots i j}^2}{4} - \frac{\sum Y_{\dots i}^2}{16} - \frac{\sum Y_{\dots j}^2}{16} + \frac{Y_{\dots}^2}{64}$
Error for var. x har.	27	by subtraction
Total	64	$\sum_g \sum_h \sum_i \sum_j n_{ghij} Y_{ghij}^2$

The above may be obtained by minimizing the residual sum of squares

$$\sum_{g=1}^k \sum_{h=1}^b \sum_{i=1}^v \sum_{j=1}^t n_{ghij} \pi_{ghij}^2$$

with respect to the various constants in the yield equation and applying standard regression theory. The normal equations obtained for the effects in the order in which they occur in the yield equation are:

$$n_{\dots} \mu + \sum_g n_{g\dots} \rho_g + \sum_h n_{\dots h} \lambda_h + \sum_i n_{\dots i} \alpha_i + \sum_g \sum_h \sum_i n_{ghi} \epsilon_{ghi}$$

$$+ \sum_j n_{...j} \beta_j + \sum_i \sum_j n_{..ij} \alpha_{ij} + \sum_h \sum_j n_{.h.j} \delta_{hj} = Y_{...}$$

$$n_{g...}(\mu + \rho_g) + \sum_h n_{gh..} \lambda_h + \sum_i n_{g.i.} \alpha_i + \sum_h \sum_i n_{ghi.} \epsilon_{ghi}$$

$$+ \sum_j n_{g..j} \beta_j + \sum_i \sum_j n_{g.ij} \alpha_{ij} + \sum_h \sum_j n_{gh.j} \delta_{hj} = Y_{g...}$$

$$n_{.h..}(\mu + \lambda_h) + \sum_g n_{gh..} \rho_g + \sum_i n_{.hi.} \alpha_i + \sum_g \sum_i n_{ghi.} \epsilon_{ghi}$$

$$+ \sum_j n_{.h.j}(\beta_j + \delta_{hj}) + \sum_i \sum_j n_{.hij} \alpha_{ij} = Y_{.h..}$$

$$n_{..i.}(\mu + \alpha_i) + \sum_g n_{g.i.} \rho_g + \sum_h n_{.hi.} \lambda_h + \sum_g \sum_h n_{ghi.} \epsilon_{ghi}$$

$$+ \sum_j n_{..ij}(\beta_j + \alpha_{ij}) + \sum_h \sum_j n_{.hij} \delta_{hj} = Y_{..i.}$$

$$n_{ghi.}(\mu + \rho_g + \lambda_h + \alpha_i + \epsilon_{ghi}) + \sum_j n_{ghij}(\beta_j + \alpha_{ij} + \delta_{hj}) = Y_{ghi.}$$

$$n_{...j}(\mu + \beta_j) + \sum_g n_{g..j} \rho_g + \sum_h n_{.h.j}(\lambda_h + \delta_{hj}) + \sum_i n_{..ij}(\alpha_i + \alpha_{ij})$$

$$+ \sum_g \sum_h \sum_i n_{ghij} \epsilon_{ghi} = Y_{...j}$$

$$n_{..ij}(\mu + \alpha_i + \beta_j + \alpha\beta_{ij}) + \sum_g n_{g \cdot ij} \rho_g + \sum_h n_{\cdot hij}(\lambda_h + \delta_{hj}) + \sum_g \sum_h n_{ghij} \epsilon_{ghi} = Y_{..ij}$$

$$n_{\cdot h \cdot j}(\mu + \lambda_h + \beta_j + \delta_{hj}) + \sum_g n_{gh \cdot j} \rho_g + \sum_i n_{\cdot hij}(\alpha_i + \alpha\beta_{ij}) + \sum_g \sum_i n_{ghij} \epsilon_{ghi} = Y_{\cdot h \cdot j}$$

Utilizing the following restraint equations:

$$\sum \rho_g = \sum \lambda_h = \sum \alpha_i = \sum \beta_j = \sum_{h \text{ or } j} \delta_{hj} = \sum_{i \text{ or } j} \alpha\beta_{ij} = \sum_{g, h, \text{ or } i} \epsilon_{ghi} = 0,$$

solutions for the effects are easily found if there are no missing observations.

These are $\hat{\mu} = \bar{y}$, $\hat{\mu} + \hat{\rho}_g = \bar{y}_{g \cdot}$, $\hat{\mu} + \hat{\lambda}_h = \bar{y}_{\cdot h}$, $\hat{\mu} + \hat{\alpha}_i = \bar{y}_{\cdot \cdot i}$, $\hat{\mu} + \hat{\rho}_g + \hat{\lambda}_h + \hat{\alpha}_i + \hat{\epsilon}_{ghi} = \bar{y}_{ghi}$, $\hat{\mu} + \hat{\beta}_j = \bar{y}_{\cdot \cdot j}$, $\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha}\hat{\beta}_{ij} = \bar{y}_{\cdot \cdot ij}$, and $\hat{\mu} + \hat{\lambda}_h + \hat{\beta}_j + \hat{\delta}_{hj} = \bar{y}_{\cdot h \cdot j}$, where these are the ordinary arithmetic means of the previously defined totals.

When there are missing observations the solution to the normal and restraint equations may become difficult depending upon the distribution of the missing observations. Instead of considering all the constants let us consider the following constants and let us note that the remaining normal equations may be obtained as sums of normal equations for these constants:

$$\mu + \rho_g + \lambda_h + \alpha_i + \epsilon_{ghi} = \epsilon_{ghi}^*$$

$$\mu + \alpha_i + \beta_j + \alpha\beta_{ij} = \alpha\beta_{ij}^*$$

$$\mu + \lambda_h + \beta_j + \delta_{hj} = \delta_{hj}^*$$

Then, the normal equations may be written as:

$$\begin{pmatrix} D_\epsilon & N_1 & N_2 \\ N_1' & D_{\alpha\beta} & N_3 \\ N_2' & N_3' & D_\delta \end{pmatrix} \begin{pmatrix} \epsilon^* \\ \alpha\beta^* \\ \delta^* \end{pmatrix} = \begin{pmatrix} Y_\epsilon \\ Y_{\alpha\beta} \\ Y_\delta \end{pmatrix}$$

where D_ϵ is a $bk \times bk$ diagonal matrix with diagonal elements $n_{ghi.}$, $D_{\alpha\beta}$ is a $tv \times tv$ diagonal matrix with diagonal elements $n_{.ij}$, D_δ is a $bt \times bt$ diagonal matrix with diagonal elements $n_{.h.j}$, $\epsilon^* = (\epsilon_{111}^* \epsilon_{212}^* \epsilon_{313}^* \epsilon_{414}^* \epsilon_{122}^* \epsilon_{223}^* \epsilon_{324}^* \epsilon_{421}^* \epsilon_{133}^* \epsilon_{234}^* \epsilon_{331}^* \epsilon_{432}^* \epsilon_{144}^* \epsilon_{241}^* \epsilon_{342}^* \epsilon_{443}^*)'$, $\alpha\beta^* = (\alpha\beta_{11}^* \alpha\beta_{12}^* \alpha\beta_{13}^* \alpha\beta_{14}^* \alpha\beta_{21}^* \alpha\beta_{22}^* \alpha\beta_{23}^* \alpha\beta_{24}^* \alpha\beta_{31}^* \alpha\beta_{32}^* \alpha\beta_{33}^* \alpha\beta_{34}^* \alpha\beta_{41}^* \alpha\beta_{42}^* \alpha\beta_{43}^* \alpha\beta_{44}^*)'$, $\delta^* = (\delta_{11}^* \delta_{12}^* \delta_{13}^* \delta_{14}^* \delta_{21}^* \delta_{22}^* \delta_{23}^* \delta_{24}^* \delta_{31}^* \delta_{32}^* \delta_{33}^* \delta_{34}^* \delta_{41}^* \delta_{42}^* \delta_{43}^* \delta_{44}^*)'$, Y_ϵ is $bk \times 1$ with elements $Y_{ghi.}$ in the same order as ϵ^* , $Y_{\alpha\beta}$ is $tv \times 1$ with elements $Y_{.ij}$ in the same order as $\alpha\beta^*$, Y_δ is $bt \times 1$ with elements $Y_{.h.j}$ in the same order as δ^* , and N_1 , N_2 , and N_3 are the corresponding coefficient matrices for the remaining constants in the normal equations. For example, for the design given above N_1 is the following $bk \times vt$ matrix:

$$\begin{bmatrix}
 (n_{111j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & (n_{212j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{313j}) & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{414j}) \\
 0 & 0 & 0 & 0 & (n_{122j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{223j}) & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{324j}) \\
 (n_{421j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{133j}) & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{234j}) \\
 (n_{331j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & (n_{432j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{144j}) \\
 (n_{241j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & (n_{342j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{443j}) & 0 & 0 & 0 & 0
 \end{bmatrix}$$

where the matrix (n_{111j}) , e.g., is a one \times t row vector of ones if there are no missing observations and has zeros substituted for the ones where there are missing observations, and similarly for the remaining row vectors (n_{ghij}) .

For the above design N_2 is a $bk \times bt = 16 \times 16$ matrix of the form:

$$\begin{bmatrix}
 (n_{111j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 (n_{212j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 (n_{313j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 (n_{414j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & (n_{122j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & (n_{223j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & (n_{324j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & (n_{421j}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{133j}) & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{234j}) & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{331j}) & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{432j}) & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{144j}) \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{241j}) \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{342j}) \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (n_{443j})
 \end{bmatrix}$$

For the above design N_3 is a $bk \times vt = 16 \times 16$ matrix of the form:

$$\begin{bmatrix}
 & & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & E & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & & & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & & F & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & & & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & & & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & G & & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & H & \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & 0
 \end{bmatrix}$$

where

$$E = \begin{bmatrix}
 n_{1111} & n_{4211} & n_{3311} & n_{2411} \\
 n_{1112} & n_{4212} & n_{3312} & n_{2412} \\
 n_{1113} & n_{4213} & n_{3313} & n_{2413} \\
 n_{1114} & n_{4214} & n_{3314} & n_{2414}
 \end{bmatrix}$$

$$F = \begin{bmatrix}
 n_{2121} & n_{1221} & n_{4321} & n_{3421} \\
 n_{2122} & n_{1222} & n_{4322} & n_{3422} \\
 n_{2123} & n_{1223} & n_{4323} & n_{3423} \\
 n_{2124} & n_{1224} & n_{4324} & n_{3424}
 \end{bmatrix}$$

$$G = \begin{bmatrix} n_{3131} & n_{2231} & n_{1331} & n_{4431} \\ n_{3132} & n_{2232} & n_{1332} & n_{4432} \\ n_{3133} & n_{2233} & n_{1333} & n_{4433} \\ n_{3134} & n_{2234} & n_{1334} & n_{4434} \end{bmatrix}$$

$$H = \begin{bmatrix} n_{4141} & n_{3241} & n_{2341} & n_{1441} \\ n_{4142} & n_{3242} & n_{2342} & n_{1442} \\ n_{4143} & n_{3243} & n_{2343} & n_{1443} \\ n_{4144} & n_{3244} & n_{2344} & n_{1444} \end{bmatrix}$$

In order to solve for the ϵ^* , $\alpha\beta^*$, and δ^* we may rewrite the normal equations as:

$$\begin{pmatrix} D_\epsilon & N_1 & N_2 \\ N'_1 & D_{\alpha\beta} & N_3 \\ N'_2 & N'_3 & D_\delta \end{pmatrix} \begin{pmatrix} \epsilon^* \\ \alpha\beta^* \\ \delta^* \end{pmatrix} = \begin{pmatrix} Y_\epsilon \\ Y_{\alpha\beta} \\ Y_\delta \end{pmatrix}$$

$$= \begin{pmatrix} D_\epsilon & N_1 & N_2 \\ 0 & D_{\alpha\beta} - N'_1 D^{-1} N_1 = B_{\alpha\beta} & N_3 - N'_1 D^{-1} N_2 = N_4 \\ 0 & N'_3 - N'_2 D^{-1} N_1 = N_4 & D_\delta - N'_2 D^{-1} N_2 = B_\delta \end{pmatrix} \begin{pmatrix} \epsilon^* \\ \alpha\beta^* \\ \delta^* \end{pmatrix} = \begin{pmatrix} Y_\epsilon \\ Y_{\alpha\beta} - N'_1 D^{-1} Y_\epsilon = X_{\alpha\beta} \\ Y_\delta - N'_2 D^{-1} Y_\epsilon = X_\delta \end{pmatrix}$$

$$= \begin{pmatrix} D_{\epsilon} & N_1 & N_2 \\ 0 & B_{\alpha\beta} & N_4 \\ 0 & 0 & B_{\delta} - N_4' B_{\alpha\beta}^{-1} N_4 \end{pmatrix} \begin{pmatrix} \epsilon^* \\ \alpha\beta^* \\ \delta^* \end{pmatrix} = \begin{pmatrix} Y_{\epsilon} \\ X_{\alpha\beta} \\ X_{\delta} - N_4' B_{\alpha\beta}^{-1} X_{\alpha\beta} \end{pmatrix}$$

The above implies that none of the diagonal elements of D_{ϵ} , $D_{\alpha\beta}$, or D_{δ} is zero.

If this is the case delete the corresponding row and column from the matrix and the corresponding constant. Also, implied in the above is that $B_{\alpha\beta}$ has an inverse but this is generally not the case and use must be made of a

restraint equation of the form $\sum_{ij} (\alpha\beta_{ij}^* - \mu) = 0$, prior to the last step above.

Then, in order to solve for the δ^* one must impose a restraint of the form

$\sum_{hj} (\delta_{hj}^* - \mu) = 0$. Restraint equations of the form listed previously must be

utilized if solutions (denoted by the constant with a caret as e.g. $\hat{\alpha}_i$) of

the $\alpha_i, \rho_g, \lambda_h, \beta_j, \epsilon_{ghi}, \alpha\beta_{ij},$ and δ_{hj} are to be obtained. After obtaining

solutions for all constants we may compute the residual sum of squares as

$$R = \sum_{ghij} \sum_{ghij} (Y_{ghij} - \hat{\mu} - \hat{\rho}_g - \hat{\lambda}_h - \hat{\alpha}_i - \hat{\epsilon}_{ghi} - \hat{\beta}_j - \hat{\alpha}\hat{\beta}_{ij} - \hat{\delta}_{hj})^2 = \sum_{ghij} \sum_{ghij} Y_{ghij}^2 - S \text{ where}$$

$$S = SS(\mu, \rho_g, \lambda_h, \alpha_i, \epsilon_{ghi}, \beta_j, \alpha\beta_{ij}, \delta_{hj}) = \hat{\mu}Y \dots + \sum \hat{\rho}_g Y_g \dots + \sum \hat{\lambda}_h Y_{\cdot h} \dots + \sum \hat{\alpha}_i Y_{\dots i}$$

$$+ \sum \sum \sum n_{ghi} \hat{\epsilon}_{ghi} Y_{ghi} + \sum \hat{\beta}_j Y_{\dots j} + \sum \sum \hat{\alpha}\hat{\beta}_{ij} Y_{\dots ij} + \sum \sum \hat{\delta}_{hj} Y_{\dots hj} \dots$$

The sum of squares due to any set of constants, say ϵ_{ghi} , eliminating all other effects, is equal to

$$S - SS(\mu, \rho_g, \lambda_h, \alpha_i, \beta_j, \alpha\beta_{ij}, \delta_{hj}) = S_1 \text{ with } f_1 \text{ degrees of freedom. The sum of}$$

squares due to two sets of constants, say α_i and ϵ_{ghi} , eliminating all other

effects is $S - SS(\mu, \rho_g, \lambda_h, \beta_j, \alpha\beta_{ij}, \delta_{hj}) = S_2$ with f_2 degrees of freedom. Then

we may utilize a general theorem from least squares theory which states that

when the null hypothesis is true $\frac{(S_2 - S_1)/(f_2 - f_1)}{S_1/f_1}$ follows Snedecor's F distribu-

tion with $f_2 - f_1$ and f_1 degrees of freedom. From this result, one may obtain F

tests for harvests, for varieties, and for the harvest x variety interaction regardless of the balanced or unbalanced nature of the design.

For the particular case at hand, the experimenter did not obtain results for harvest number four. This means that additional restraint equations of the form $\alpha_{4j}=0$, $\alpha\beta_{4j}=0$, $\epsilon_{gh4}=0$, and $n_{gh4j}=0$ for all cases where n_{gh4j} was one with no missing observations. For this case the solutions to the normal equations become:

$$\hat{\mu} = \bar{y}, \hat{\alpha}_i = \bar{y}_{..i} - \bar{y}, \hat{\beta} = \bar{y}_{...j} - \bar{y},$$

$$\hat{\alpha\beta}_{ij} = \bar{y}_{..ij} - \bar{y}_{..i} - \bar{y}_{...j} + \bar{y},$$

solutions for ρ_g are obtained from the equations

$$\begin{pmatrix} 32/3 & 0 & 0 & 0 \\ 0 & 32/3 & 0 & 0 \\ 0 & 0 & 32/3 & 0 \\ 0 & 0 & 0 & 32/3 \end{pmatrix} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\rho}_3 \\ \hat{\rho}_4 \end{pmatrix} = \begin{pmatrix} Y_{1...} - (Y_{....} - Y_{.4..})/3 = Q_{r1} \\ Y_{2...} - (Y_{....} - Y_{.3..})/3 = Q_{r2} \\ Y_{3...} - (Y_{....} - Y_{.2..})/3 = Q_{r3} \\ Y_{4...} - (Y_{....} - Y_{.1..})/3 = Q_{r4} \end{pmatrix},$$

and the solutions for λ_h are obtained from the equations

$$\begin{pmatrix} 32/3 & 0 & 0 & 0 \\ 0 & 32/3 & 0 & 0 \\ 0 & 0 & 32/3 & 0 \\ 0 & 0 & 0 & 32/3 \end{pmatrix} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \hat{\lambda}_3 \\ \hat{\lambda}_4 \end{pmatrix} = \begin{pmatrix} Y_{.1...} - (Y_{....} - Y_{4....})/3 = Q_{c1} \\ Y_{.2...} - (Y_{....} - Y_{3....})/3 = Q_{c2} \\ Y_{.3...} - (Y_{....} - Y_{2....})/3 = Q_{c3} \\ Y_{.4...} - (Y_{....} - Y_{1....})/3 = Q_{c4} \end{pmatrix}$$

$$\hat{\epsilon}_{ghi} = \bar{y}_{ghi} - \bar{y} - \hat{\rho}_g - \hat{\lambda}_h - \hat{\alpha}_i, \text{ and } \hat{\delta}_{hj} = \frac{1}{3}(Y_{.h.j} - \frac{1}{4}Y_{.h..}) - \hat{\beta}_j = \bar{y}_{.h.j} - \bar{y}_{.h..} - \bar{y}_{...j} + \bar{y}.$$

For the case where harvest number 4 is missing all effects are orthogonal to each other except rows and columns. The analysis of variance table is:

<u>Source of variation</u>	<u>d.f.</u>	<u>Sum of squares</u>
Total	48	$\sum_{g=1}^4 \sum_{h=1}^4 \sum_{i=1}^3 \sum_{j=1}^4 n_{ghij} Y_{ghij}^2$
Correction for the mean	1	$Y_{...}^2/48$
Row (ignoring column)	3	$\sum \frac{Y_{g...}^2}{12} - \frac{Y_{...}^2}{48} = R$
Column (eliminating row)	3	$\sum \hat{\lambda}_h \hat{Q}_{c.h} = C$
Harvest	2	$\sum \frac{Y_{.h..}^2}{16} - \frac{Y_{...}^2}{48} = H$
Error for harvest	3	$\sum_g \sum_i \frac{Y_{g.hi.}^2}{4} - R - C - H - \frac{Y_{...}^2}{48}$
Variety	3	$\sum \frac{Y_{...i}^2}{12} - \frac{Y_{...}^2}{48}$
Error for variety	9	$\sum \sum \frac{Y_{.h..i}^2}{3} - \sum \frac{Y_{.h..}^2}{12} - \sum \frac{Y_{...i}^2}{12} + \frac{Y_{...}^2}{48}$
Harvest x variety	6	$\sum \sum \frac{Y_{.h..i}^2}{4} - \sum \frac{Y_{.h..}^2}{16} - \sum \frac{Y_{...i}^2}{12} + \frac{Y_{...}^2}{48}$
Error for harvest x variety	18	subtraction

With harvest 4 data deleted the analysis of variance and solution of effects remain simple. With other missing plot situations the results are likely to become more complicated and the various effects are likely to be non-orthogonal, thus complicating the analysis. The procedure described above holds for all unbalanced situations. Also, deletion of the δ_{hj} from the above yield equations gives the yield equation for the split-plot design. Addition of other constants in the yield equation results in the split-split plot design, split block design with split plots, etc. Missing observations in all these designs may be handled as described above.

Alternatively, one may include dummy independent variates in regression where the X_{fghij} values take on one (or -1) when the yield is missing and zero otherwise. Then, a multiple covariance analysis of the yield Y_{ghij} on the independent variates X_{fghij} is computed. The solutions for the constants and the sums of squares in the analysis of variance adjusted for the covariates are identical to those obtained from regression theory.