# A Procedure for Estimating Genetic Advance

## in a Multi-Site Trial

W. T. Federer

## Abstract

For a maize experiment involving 196 entries (including 6 checks) designed as a simple lattice and carried out over six sites, it was desired to develop a procedure to obtain entry means adjusted for interblock information and for multiple covariates at each site. An estimate of a component of variance for the 190 non-check entries at each site was desired. Also, estimates of entry × site and entry components of variance over sites were desired. Then, reallocation of resources would be studied to determine how to maximize genetic advance.

## 1. Introduction

An experiment of 190 maize entries plus six maize checks was designed as a simple (double) lattice design and the experiment was conducted at six locations (sites). In addition to the yield of grain (unshelled) of maize and other characteristics, the number of plants (stand) in each plot (experimental unit) was obtained. It was thought that number of plants in the plot, number of plants in the two adjacent plots, and perhaps number of plants per plot squared should be used as four covariates. However, only the two covariates, number of plants per plot and the sum of the number of plants in the two adjacent plots, may be all that is necessary to account for stand variations and competition between adjacent plots.

The method for analyzing data from a simple lattice experiment design with a covariate is described by Federer (1967), Sections XI–3.1, XI–8, XVI–7, and XVI–12.3. Following the outline of the analysis given there, we extend it to multiple covariance. In addition, it is desired to estimate a

variance component for the 190 entries in the experiment and to use this in calculating genetic advance as described by Federer (1951), Rojas and Sprague (1952), and Sprague and Federer (1951).

In the following section, an analysis of covariance with multiple covariates is described for a simple lattice design. The results are illustrated with a numerical example. Anyone using computer software such as SAS should first try to obtain the results given in the numerical example. Computer programs do not always give the desired results, despite the claims as advertised. A list of references where this has been discovered is available upon request.

In the third section, the expected values for the various mean squares from such an experiment as described above is obtained and is illustrated with a numerical example. Here again it is recommended that any computer program for obtaining coefficients for variance components be checked against the example. This is a necessary but not sufficient check on the correctness of a computer program.

In the fourth section, we show how to combine experiments of the nature following a procedure suggested by W. G. Cochran and given in Cochran and Cox (1957), Sections 14.32 and 14.4. The entry means adjusted for interblock information and covariates are then used to study entry by environment interactions as described in Basford, McCulloch, and Murty (1992).

In the last section, a discussion of other procedures is given. The interblock regressions could be different from the intrablock regression. This would be similar to the split plot situation where the error (a) and error (b) regressions differ (Federer and Meredith, 1991). Also, since the 190 entries represent a sample from a population, account needs to be taken of the distributional properties of entries (random effects). Some sort of shrinkage estimator such as BLUP (best linear unbiased predictor) may be appropriate.

## 2. Covariance Analysis for a Simple Lattice with Multiple Covariates

For the experiment described above, let us use the following response model for the $g^{th}$ entry in the $ef^{th}$ incomplete block in the $e^{th}$ complete block:

$$Y_{efg} = \mu + \rho_e + \pi_{ef} + \tau_g + \epsilon_{efg} + \sum_{d=1}^{p} \beta_d\left(X_{defg} - \overline{x}_{d\ldots}\right), \tag{1}$$

where $\mu$ is an effect common to every observation (response), $\rho_e$ is the effect associated with the $e^{th}$ complete block, $\pi_{ef}$ is the effect associated with the $f^{th}$ incomplete block within the $e^{th}$ complete block, $\tau_g$ is the effect associated with the $g^{th}$ entry, $\epsilon_{efg}$ is a random error component which is identically and independently distributed with mean zero and variance $\sigma_\epsilon^2$, $\beta_d$ is the partial regression coefficient associated with covariate $X_{defg}$, and $\overline{x}_{d\ldots}$ is the arithmetic mean of the $d^{th}$ covariate. The subscript $e = 1$, $2 = r$ denotes the number of complete blocks here, $f = 1$, $2$, $\cdots$, $14 = k =$ the number of incomplete blocks in the $e^{th}$ complete blocks, and $g = 1$, $2$, $\cdots$, $196 = k^2 =$ the number of entries. It is convenient to number the entries as for a two-factor factorial with $k$ levels of each factor. Then, let the levels of one factor be confounded with incomplete blocks in one of the two arrangements; also, let the levels of the other factor be completely confounded with incomplete blocks in the second confounding arrangement. This gives one-half intra-incomplete block (intrablock) information on the two pseudo-main effects and full intrablock information on the pseudo-interaction effects. This results in an intrablock efficiency factor of $k/(k+1) = 14/15 = 93.3\%$ for this experiment. When interblock information is recovered the efficiency factor approaches unity (see Federer and Speed, 1987), and depends upon the ratio of $\sigma_\epsilon^2 / \sigma_\pi^2$ where the $\pi_{ef}$ are identically and independently distributed with mean zero and variance $\sigma_\pi^2$.

The sums of squares and cross products of the response Y and the p covariates, $X_1$, $\cdots$, $X_p$, are computed in the usual manner for a randomized complete block design for the first five sources of variation given in Table 1 (these are listed in matrix form to conserve space). If the two pseudo-factors are $A_i$ and $B_j$, $i$, $j = 0$, $1$, $\cdots$, $k-1$, then the incomplete block (eliminating entry effects) with $2(k-1)$ degrees of freedom is computed as

$$B_{yy} = \sum_{i=0}^{k-1}\left(A_{ci} - A_{ui}\right)^2 \Big/ 2k + \sum_{j=0}^{k-1}\left(B_{cj} - B_{uj}\right)^2 \Big/ 2k - \frac{\left(Y_1.. - Y_2..\right)^2}{k^2}, \tag{2}$$

Table 1. Analysis of covariance for a simple lattice design
with p covariates and intrablock regression.

| Source of variation | Degrees of freedom | Sums of squares and products |
|---|---|---|
| Total | $2k^2$ | $$\begin{bmatrix} YY & YX_1 & YX_2 & \cdots & YX_p \\ YX_1 & X_1X_1 & X_1X_2 & \cdots & X_1X_p \\ \vdots & & & & \\ YX_p & X_1X_p & X_2X_p & \cdots & X_pX_p \end{bmatrix}$$ |
| Correction for mean | 1 | $$\begin{bmatrix} Y...Y... & Y...X_1... & \cdots & Y...X_p... \\ \vdots & & & \\ Y...X_{p\cdot}... & X_1...X_{p\cdot}... & \cdots & X_{p\cdot}..X_{p\cdot}... \end{bmatrix}$$ |
| Complete block = R | 1 | $$\begin{bmatrix} R_{yy} & R_{y1} & \cdots & R_{yp} \\ \vdots & & & \\ R_{yp} & R_{1p} & \cdots & R_{pp} \end{bmatrix}$$ |
| Entry (ignoring incomplete block) = V | $k^2-1$ | $$\begin{bmatrix} V_{yy} & V_{y1} & \cdots & V_{yp} \\ \vdots & & & \\ V_{yp} & V_{1p} & \cdots & V_{pp} \end{bmatrix}$$ |
| R × V | | $$\begin{bmatrix} RV_{yy} & RV_{y1} & \cdots & RV_{yp} \\ \vdots & & & \\ RV_{yp} & RV_{1p} & \cdots & RV_{pp} \end{bmatrix}$$ |
| Incomplete block (eliminating entry) | $2(k-1)$ | $$\begin{bmatrix} B_{yy} & B_{y1} & \cdots & B_{yp} \\ \vdots & & & \\ B_{yp} & B_{1p} & \cdots & B_{pp} \end{bmatrix}$$ |
| Intrablock | $(k-1)^2$ | $$\begin{bmatrix} E_{yy} & E_{y1} & \cdots & E_{yp} \\ \vdots & & & \\ E_{yp} & E_{1p} & \cdots & E_{pp} \end{bmatrix}$$ |

Table 1 (continued)

**Adjusted sums of squares**

| Source of variation | Degrees of freedom | Sums of squares and products |
|---|---|---|
| Incomplete block (eliminating entry and regression) | $2(k-1)$ | $\left(1-R_{RV}^2\right)RV_{yy} - \left(1-R_E^2\right)E_{yy} = B'_{yy}$ |
| Intrablock | $(k-1)^2-p$ | $\left(1-R_E^2\right)E_{yy} = E'_{yy}$ |
| Entry (adjusted for incomplete block) | $k^2-1$ | $\begin{bmatrix} T_{yy} & T_{y1} & \cdots & T_{yp} \\ \vdots & & & \\ T_{yp} & T_{1p} & \cdots & T_{pp} \end{bmatrix}$ |
| Entry (adjusted for incomplete block and regression) | $(k^2-1)$ | $\left(1-R_{T+E}^2\right)(T_{yy}+E_{yy}) - E'_{yy} = T'_{yy}$ |
| Incomplete block (eliminating entry) regression | $p$ | $R_B^2\, B_{yy}$ |
| Deviations from block regression | $2(k-1)-p$ | $\left(1-R_B^2\right)B_{yy}$ |
| Intrablock regression | $p$ | $R_E^2 E_{yy}$ |

where $A_{ci}$ is the block total corresponding to the $i^{th}$ level of factor A, $A_{ui}$ is the $i^{th}$ level of factor A in the complete block where it is unconfounded with incomplete blocks, $B_{cj}$ is the block total corresponding to the $j^{th}$ level of factor B in the complete block where it is confounded with incomplete blocks, $B_{uj}$ is the $j^{th}$ level of factor B in the complete block where it is unconfounded with incomplete blocks, and $Y_{e..}$ is the $e^{th}$ complete block total. When k is a prime or prime power, there are $k+1$ main effects and interactions, and sums of squares of levels among these effects, where they are unconfounded with incomplete blocks, results in the intrablock error sum of squares. For any k, the residuals could be computed, squared, and summed to obtain the intrablock error sum of squares. However, it is usually obtained by subtracting the sum of squares or products in (2) from the R × E sum of squares or products. Note that formula (2) is useful for computing products simply by writing $(A_{ci}-A_{ui})^2$ as $(A_{ci}-A_{ui})(A_{ci}-A_{ui})$ and letting one of the quantities in parenthesis be for one variate, say $X_1$, and letting the other quantity in parenthesis be for another variate, say $X_p$. The correction term $(Y_1..-Y_2..)^2\big/k^2$ would be replaced by $(X_{11}..-X_{12}..)(X_{p1}..-X_{p2}..)\big/k^2$, taking into account the sign of the difference of the quantity in the parenthesis.

The sums of squares and products in the last matrix in Table 1 are used to compute $R_E^2$ and the estimates of $\beta_d$, the intrablock regressions. The matrix of sums of squares and products associated with R × V is used to compute $R_{RV}^2 = R_{B+E}^2$. Likewise, the sums of squares and products for entry (eliminating incomplete blocks) are used to compute $R_{T+E}^2$. To obtain $R_B^2$, use the sums of squares and products associated with the incomplete block (eliminating entry) source of variation. The sums of squares and products for entry (eliminating incomplete block effects) may be computed in various ways. For example,

$$\text{Entry (ignoring blocks)} + \text{Block (eliminating entry effects)}$$
$$- \text{Block (ignoring entry effects)} = \text{Entry (eliminating block effects)} \tag{3}$$

or solutions for the $\tau_g$ in matrix form under the restraint $\sum_{g=1}^{k^2} \hat{\tau}_g = 0$, e.g., are given by

$$\hat{\underline{\tau}} = \left[2\mathbf{I} - \frac{1}{k}\mathbf{NN}' + c\mathbf{J}\right]^{-1}\left[\underline{Y}_V - \mathbf{N}'\underline{Y}_B\big/k\right] \tag{4}$$

and the sum of squares in (3) is equal to

$$\hat{\underline{\tau}}'\left[\underline{Y}_V - N'\underline{Y}_B/k\right] = \underline{\tau}'\underline{Q}_V = \sum_{g=1}^{k^2} \hat{\tau}_g\, Q_{..g}\,, \tag{5}$$

where $\hat{\underline{\tau}}$ is a $k^2 \times 1$ vector of $\hat{\tau}_g$, $\underline{Y}_V$ is a $k^2 \times 1$ vector of entry totals, $\underline{Y}_B$ is a $2k \times 1$ vector of incomplete block totals, $N$ is a $2k \times k^2$ matrix of zeros and ones to indicate when entry g occurs in block ef, $n_{efg} = 0, 1$ are the elements of $N$, $I$ is a $k^2 \times k^2$ identity matrix, 2 is the number of replicates of each entry, $J$ is a $k^2 \times k^2$ matrix whose elements are all ones $\left(\text{i.e., } \sum_{g=1}^{k^2} \hat{\tau}_g = 0\right)$, and c is a constant which makes as many zeros in $NN'/k+cJ$ as possible under the restraint $\sum \hat{\tau}_g = 0$. The inverse of $\left[2I - NN'/k+cJ\right]$ times $\sigma_E^2$ is the intrablock variance-covariance matrix for $\hat{\underline{\tau}}$.

To illustrate the above and to provide a numerical example as a necessary check on the calculations from a software package such as SAS, we present the following example.

Example 1. The example is the one given by Federer (1967), Example XI-1, and two covariates $X_1$ and $X_2$ have been added. The numerical values are given in Table 2. The factor B is confounded with incomplete blocks in complete block 1, and factor A in 2. The unconfounded levels of effects versus confounded levels of effects are given as $A_{.i} - 2A_{2i}$ and $B_{.j} - 2B_{1j}$, as this computational form eliminates the necessity of computing the $A_{1i}$ and $B_{2j}$ values. The computations for the R × E sums of squares and products are given in Table 3. Likewise, the computations for incomplete block (eliminating entry effects) are also given in the table. The intrablock sums of squares and products were obtained by subtraction. Note that in computing the randomized complete block residuals $\left(Y_{efgh} - \bar{Y}_{e...} - \bar{Y}_{..gh} + \bar{Y}_{....}\right)$ these were all multiplied by the number of plots in the experiment, i.e., $2k^2$, and they were computed as $2k^2 Y_{efgh} - 2Y_{..gh} - k^2 Y_{e...} + Y_{....}$ to eliminate rounding errors. $Y_{..gh}$ is the $g^{th}$ entry total, $Y_{e...}$ is the $e^{th}$ complete block total, and $Y_{....}$ is the grand total of the $2k^2$ plots.

The various sums of squares and products for the data in Table 2 are given in Table 3. The entries (eliminating incomplete blocks) sum of squares for Y may be obtained by first computing the $Q_{..gh}$ and $\hat{\mu} + \hat{\tau}_{gh}$ (or $\hat{\tau}_{gh}$) values as:

Table 2. Responses $Y_{efgh}$ for nine entries (gh = 00, 01, 02, 20, 11, 13, 20, 21, and 22) in a simple lattice design with two covariates.

| Block 11 Entry(gh) | $Y_{11gh}$ | $X_{111gh}$ | $X_{211gh}$ | Block 12 Entry | $Y_{12gh}$ | $X_{112gh}$ | $X_{212gh}$ | Block 13 Entry | $Y_{13gh}$ | $X_{113gh}$ | $X_{213gh}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 00 | 8 | 3 | 9 | 02 | 3 | 1 | 1 | 21 | 3 | 2 | 1 |
| 20 | 5 | 2 | 4 | 12 | 2 | 1 | 1 | 11 | 7 | 3 | 4 |
| 10 | 3 | 1 | 1 | 22 | 6 | 3 | 4 | 01 | 3 | 1 | 1 |
| $B_{10}$ | 16 | 6 | 14 | $B_{12}$ | 11 | 5 | 6 | $B_{11}$ | 13 | 6 | 6 |

| Block 21 Entry(gh) | $Y_{21gh}$ | $X_{121gh}$ | $X_{221gh}$ | Block 22 Entry | $Y_{22gh}$ | $X_{122gh}$ | $X_{222gh}$ | Block 23 Entry | $Y_{23gh}$ | $X_{123gh}$ | $X_{223gh}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 21 | 2 | 1 | 1 | 10 | 3 | 1 | 4 | 01 | 2 | 1 | 4 |
| 20 | 2 | 1 | 1 | 11 | 3 | 2 | 1 | 02 | 4 | 1 | 1 |
| 22 | 7 | 3 | 9 | 12 | 3 | 1 | 1 | 00 | 6 | 3 | 9 |
| $A_{22}$ | 11 | 5 | 11 | $A_{21}$ | 9 | 4 | 6 | $A_{20}$ | 12 | 5 | 14 |

Entry totals (Y)

| | | | | | | | $A._g$ | $A_{2g}$ | $A._g-2A_{2g}$ |
|---|---|---|---|---|---|---|---|---|---|
| 00 | 14 | 01 | 5 | 02 | 7 | | 26 | 12 | 2 |
| 10 | 6 | 11 | 10 | 12 | 5 | | 21 | 9 | 3 |
| 20 | 7 | 21 | 5 | 22 | 13 | | 25 | 11 | 3 |
| $B._h$ | 27 | | 20 | | 25 | | 72 | 32 | 8 |
| $B_{1h}$ | 16 | | 13 | | 11 | | 40 | 72 | |
| $B._h-2B_{1h}$ | -5 | | -6 | | 3 | | -8 | | 0 |

Entry totals ($X_1$)

| | | | | | | | $A._g$ | $A_{2g}$ | $A._g-2A_{2g}$ |
|---|---|---|---|---|---|---|---|---|---|
| 00 | 6 | 01 | 2 | 02 | 2 | | 10 | 5 | 0 |
| 10 | 2 | 11 | 5 | 12 | 2 | | 9 | 4 | 1 |
| 20 | 3 | 21 | 3 | 22 | 6 | | 12 | 5 | 2 |
| $B._h$ | 11 | | 10 | | 10 | | 31 | 14 | 3 |
| $B_{1h}$ | 6 | | 6 | | 5 | | 17 | 31 | |
| $B._h-2B_{1h}$ | -1 | | -2 | | 9 | | -3 | | 0 |

Entry totals ($X_2$)

| | | | | | | | $A._g$ | $A_{2g}$ | $A._g-2A_{2g}$ |
|---|---|---|---|---|---|---|---|---|---|
| 00 | 18 | 01 | 5 | 02 | 2 | | 25 | 14 | -3 |
| 10 | 5 | 11 | 5 | 12 | 2 | | 12 | 6 | 0 |
| 20 | 5 | 21 | 2 | 22 | 13 | | 20 | 11 | -2 |
| $B._h$ | 28 | | 12 | | 17 | | 57 | 31 | -5 |
| $B_{1h}$ | 14 | | 6 | | 6 | | 26 | 57 | |
| $B._h-2B_{1h}$ | 0 | | 0 | | 5 | | 5 | | 0 |

Table 2 (Continued)

| Totals | Block 1 | | | Block 2 | | | Grand | | |
|---|---|---|---|---|---|---|---|---|---|
| | $Y_1\ldots$ | $X_{11}\ldots$ | $X_{21}\ldots$ | $Y_2\ldots$ | $X_{12}\ldots$ | $X_{22}\ldots$ | $Y\ldots$ | $X_1\ldots$ | $X_2\ldots$ |
| | 40 | 17 | 26 | 32 | 14 | 31 | 72 | 31 | 57 |

18 (Residuals)

| Block | 00 | 01 | 02 | 10 | 11 | 12 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Y | | | | |
| 1 | 10 | 1 | −17 | −8 | 28 | −17 | 19 | 1 | −17 |
| 2 | −10 | −1 | 17 | 8 | −28 | 17 | −19 | −1 | 17 |
| | | | | | $X_1$ | | | | |
| 1 | −3 | −3 | −3 | −3 | 6 | −3 | 6 | 6 | −3 |
| 2 | 3 | 3 | 3 | 3 | −6 | 3 | −6 | −6 | 3 |
| | | | | | $X_2$ | | | | |
| 1 | 5 | −22 | 5 | −22 | 32 | 5 | 32 | 5 | −40 |
| 2 | −5 | 22 | −5 | 22 | −32 | −5 | −32 | −5 | 40 |

R × E sums of squares and products:

$$\Sigma(18\,\text{res})^2(Y)\Big/18^2 = 4356/324 = 13.4444$$

$$\Sigma(18\,\text{res})^2(X_1)\Big/18^2 = 324/324 = 1.0000$$

$$\Sigma(18\,\text{res})^2(X_2)\Big/18^2 = 9432/324 = 29.1111$$

$$\Sigma(18\,\text{res}\,Y)(18\,\text{res}\,X_1)\Big/18^2 = 864/324 = 2.6667$$

$$\Sigma(18\,\text{res}\,Y)(18\,\text{res}\,X_2)\Big/18^2 = 4446/324 = 13.7222$$

$$\Sigma(18\,\text{res}\,X_1)(18\,\text{res}\,X_2)\Big/18^2 = 1242/324 = 3.8333$$

Incomplete block (eliminating entry) sums of squares and products:

Y: $[(-5)^2+(-6)^2+3^2]\Big/6-(-8)^2\Big/18+[2^2+3^2+3^2]\Big/6-8^2\Big/18 = 8.2222$

$YX_1$: $[(-5)(-1)+(-6)(-2)+3(0)]\Big/6-(-8)(-3)\Big/18+[2(0)+3(1)+3(2)]\Big/6-8(3)\Big/18 = 1.6667$

$YX_2$: $[(-5)(0)+(-6)(0)+3(5)]\Big/6-(-8)(5)\Big/18+[2(-3)+3(0)+3(-2)]\Big/6-8(-5)\Big/18 = 4.9444$

$X_1$: $[(-1)^2+(-2)^2+0^2]\Big/6-(-3)^2\Big/18+[0^2+1^2+2^2]\Big/6-3^2\Big/18 = 0.6667$

$X_1X_2$: $[(-1)(0)+(-2)(0)+0(5)]\Big/6-(-3)(5)\Big/18+[0(-3)+1(0)+2(-2)]\Big/6-3(-5)\Big/18 = 1.000$

$X_2$: $[0^2+0^2+5^2]\Big/6-5^2\Big/18+[(-3)^2+0^2+(-2)^2]\Big/6-(-5)^2\Big/18 = 3.5556$

Table 3. Sums of squares and cross products for Y, $X_1$, and $X_2$

| Source of variation | Degrees of freedom | Sums of squares and products | | | | | |
|---|---|---|---|---|---|---|---|
| | | YY | $YX_1$ | $YX_2$ | $X_1X_1$ | $X_1X_2$ | $X_2X_2$ |
| Total | 18 | 354 | 151 | 309 | 67 | 133 | 333 |
| Correction for mean | 1 | 288 | 124 | 228 | 53.39 | 98.17 | 180.50 |
| Complete blocks = R | 1 | 3.56 | 1.33 | −2.22 | 0.50 | −0.84 | 1.39 |
| Entries = E | 8 | 49.00 | 23.00 | 69.50 | 12.11 | 31.83 | 122.00 |
| R × E | 8 | 13.44 | 2.67 | 13.72 | 1.00 | 3.84 | 29.11 |
| Incomplete block (eliminating entries) | 4 | 8.22 | 1.67 | 4.94 | 0.67 | 1.00 | 3.56 |
| Intrablock | 4 | 5.22 | 1.00 | 8.78 | 0.33 | 2.84 | 25.55 |
| Entries (eliminating incomplete blocks) | 8 | 51.4444 | 23.3333 | 63.2222 | 12.3333 | 30.5000 | 100.4444 |
| Incomplete blocks (ignoring entries) | 4 | 5.7778 | 1.3333 | 11.2222 | 0.4444 | 2.3333 | 25.1111 |

$$Q_{\cdot\cdot 00} = 14-(16+12)\big/3 = 14/3 \qquad Q_{\cdot\cdot 21} = 5-(13+11)\big/3 = -9/3$$

$$Q_{\cdot\cdot 10} = 6-(16+9)\big/3 = -7/3 \qquad Q_{\cdot\cdot 02} = 7-(11+12)\big/3 = -2/3$$

$$Q_{\cdot\cdot 20} = 7-(16+11)\big/3 = -6/3 \qquad Q_{\cdot\cdot 12} = 5-(11+9)\big/3 = -5/3$$

$$Q_{\cdot\cdot 01} = 5-(13+12)\big/3 = -10/3 \qquad Q_{\cdot\cdot 22} = 13-(11+11)\big/3 = 17/3$$

$$Q_{\cdot\cdot 11} = 10-(13+9)\big/3 = 8/3$$

$$\hat{\mu}+\hat{\tau}_{00} = 39/6 \qquad \hat{\mu}+\hat{\tau}_{01} = 11/6 \qquad \hat{\mu}+\hat{\tau}_{02} = 26/6$$

$$\hat{\mu}+\hat{\tau}_{10} = 16/6 \qquad \hat{\mu}+\hat{\tau}_{11} = 27/6 \qquad \hat{\mu}+\hat{\tau}_{12} = 21/6$$

$$\hat{\mu}+\hat{\tau}_{20} = 19/6 \qquad \hat{\mu}+\hat{\tau}_{21} = 12/6 \qquad \hat{\mu}+\hat{\tau}_{22} = 45/6$$

Note that these are intrablock solutions for entry means. Then,

$$\sum_{gh} \hat{\tau}_{gh} Q_{\cdot\cdot gh} = \sum_{gh}\left(\hat{\mu}+\hat{\tau}_{gh}\right)Q_{\cdot\cdot gh} = 51.4444 \ .$$

Note that $\hat{\mu}\Sigma Q_{\cdot\cdot gh} = 0$. Alternatively, this sum of squares may be computed from equation (3) as:

$$49 + 8.222222 - 5.777778 = 51.444444 \ ,$$

where 5.777778 is computed as:

$$\tfrac{1}{3}\left(16^2+11^2+13^2+11^2+9^2+12^2\right)-\tfrac{1}{9}\left(40^2+32^2\right)$$

and is the incomplete block within complete block (ignoring entry effects) sum of squares. The sum of products for Y and X, for blocks (ignoring entry) is

$$\tfrac{1}{3}\left[16(6)+11(5)+13(6)+11(5)+9(4)+12(5)\right]-\tfrac{1}{9}\left[40(17)+32(14)\right] = 1.333334 \ .$$

Then, the sum of products $YX_1$ for entry (eliminating blocks) is computed as

$$23 + 1.666667 - 1.333334 = 23.333333 \ .$$

The remaining sums of squares and products in Table 3 are computed in a similar manner.

The analysis of covariance given in Table 4 involves computation of several multiple correlation coefficients squared, $R^2$. Since there are only two covariates, the following computational form is convenient:

$$\left(R^2_{y12} = r^2_{y1} + r^2_{y2} - 2r_{12}r_{y1}r_{y2}\right)\big/\left(1-r^2_{12}\right), \tag{6}$$

where $r_{y1}$ is the correlation coefficients for Y and $X_1$, $r_{y2}$ is the correlation coefficient for Y and $X_2$, and $r_{12}$ is the correlation between $X_1$ and $X_2$. Another general form is

Table 4. Analysis of covariance for response Y and covariates $X_1$ and $X_2$.

| Source of variation | Degrees of freedom | Sum of squares | Mean square |
|---|---|---|---|
| Total | 18 | 354 | — |
| Correction for mean | 1 | 288 | — |
| Complete blocks = R | 1 | 3.5556 | — |
| Entries (ignoring incomplete blocks) = E | 8 | 49.0000 | — |
| R × E | 8 | 13.4444 | — |
| Incomplete blocks (eliminating entries) | 4 | 8.2222 | 2.0556 |
| Regression | 2 | 7.0736 | 3.5368 |
| Deviations from regular | 2 | 1.1486 | 0.5743 |
| Intrablock | 4 | 5.2222 | 1.3056 |
| Regression | 2 | 3.0524 | 1.5262 |
| Deviations from regular | 2 | 2.1698 | 1.0849 |
| Incomplete block (eliminating entry and regression) | 4 | 3.3138 | 0.8285 |
| Entry (eliminating incomplete block and regression) | 8 | 6.0922 | 0.7615 |

$R_E^2(\text{intrablock}) = 0.5845$

$R_B^2[\text{incomplete blocks (eliminating entries)}] = 0.8603$

$R_{B+E}^2[\text{intrablock} + \text{incomplete block (eliminating entries)}] = 0.5921$

$R_{T+E}^2[\text{intrablock} + \text{entries (eliminating blocks)}] = 0.8542$

$$R^2 = \sum_{d=1}^{P} r_{yd} b'_{yd} , \tag{7}$$

where $b'_{yd}$ is the standard partial regression coefficient for $X_d$ (see, e.g., Snedecor and Cochran, 1980, p. 357). Using (6) and the intrablock error line in Table 3,

$$R_E^2 = \frac{(0.574468 + 0.577335 - 1.118132)}{0.057608} = 0.5845 .$$

Using (7),

$$R_E^2 = 0.757937(0.3533223) + 0.759826(0.4177494) = 0.5845 .$$

In the above, $b_{y1} = 1.398493$ and $b_{y2} = 0.188430$. Then $b'_{y1} = 1.398493(0.333333/5.222222)^{\frac{1}{2}} = 0.3533223$, and $b'_{y2} = 0.188430(25.555556/5.222222)^{\frac{1}{2}} = 0.4177494$.

The incomplete block (eliminating entry) sums of squares adjusted for regression is computed as follows. The sums of squares and products for blocks (eliminating entry) and for intrablock are added together; this gives the sums of squares and products in the $R \times E$ line in Table 3. The multiple correlation squared on this line is computed as

$$R_{RV}^2 = R_{B+E}^2 = \left[0.7272727^2 + 0.6936232^2 - 2(0.7104724)(0.7272727)(0.6936232)\right] \Big/$$

$$\left(1 - 0.7104724^2\right) = 0.5921 ,$$

where $r_{B+Ey1} = 0.7272727$, $r_{B+Ey2} = 0.6936232$, and $3.833333 \Big/ [1(29.111111)]^{\frac{1}{2}} = 0.7104724$. Then $\left(1 - R_{B+E}^2\right) RV_{yy} = (1 - 0.5921)(13.4444) = 5.483626$ and $5.483626 - \left(1 - R_E^2\right) E_{yy} = 5.483626 - 2.1698 = 3.3138$, which is the incomplete block (eliminating entry) sums of squares adjusted for intrablock error regression.

Proceeding in a similar manner, we obtain the entry (eliminating blocks) sum of squares adjusted for intrablock error regression. The multiple correlation on the entry plus error sums of squares and products is

$$R_{T+E}^2 = \left[0.9082525^2 + 0.08520859^2 - 2(0.8343770)(0.9082525)(0.8520859)\right] \Big/$$

$$\left(1 - 0.8343770^2\right) = 0.8542 ,$$

where $r_{T+Ey1} = (23.0000 + 1.0000) \Big/ [(51.444444 + 5.222222)(12.333333 + 0.333333)]^{\frac{1}{2}} = 0.9082525$, $r_{T+Ey2} = (63.222222 + 8.777778) \Big/ [(51.444444 + 5.222222)(100.444444 + 25.555556)]^{\frac{1}{2}} = 0.8520859$, and $r_{T+E12} = 0.8343770$. Then, $(1 - 0.8542)(51.444444 + 5.222222) - 2.1698 = 8.2620 - 2.1698 = 6.0922$.

The multiple correlation coefficient for block means is computed from the incomplete block (elim. entry) line in Table 3. $R_B^2 = \left[0.7118685^2 + 0.9144694^2 - 2(0.6495191)(0.7118685)(0.9144694)\right]$ $\Big/(1 - 0.6495191^2) = 0.8603$. Then $R_B^2 B_{yy} = 0.8603(8.222222) = 7.0736$, and $\left(1 - R_B^2\right)B_{yy} = (1 - 0.8603)(8.222222) = 1.1486$. Note that $R_B^2$ is larger than $R_E^2$, which may indicate that the incomplete block (eliminating entry) regression differs from the intrablock error regression. This fact would need to be taken into account when adjusting the entry means for regression (see Federer, 1967, Federer and Meredith, 1992, and Section 5 of this paper).

The next step in the analysis of a set of data as described above is to obtain the adjusted means and variances of differences among the adjusted means. The equation for adjusted means for entry gh is:

$$\bar{y}_{\cdot\cdot gh}(\text{adj.}) = \frac{1}{(r=2)}\left\{Y_{\cdot\cdot g} - \mu'\left[\left(A_{\cdot g} - A_{ug}\right)_y + (B_{\cdot h} - B_{uh})_y\right.\right.$$

$$\left.\left. - \sum_{d=1}^{p}\hat{\beta}_d\left\{\left(A_{\cdot g} - A_{ug}\right)_d + (B_{\cdot h} - B_{uh})\right\}\right]\right\}, \tag{8}$$

where $\mu' = (w - w')\big/k(w + w')$, $w = 1\big/E'_{yy} = 1\big/1.0849 = 0.9217$, $w' = 1\big/\left[2(0.8285) - 1.0849\right] = 1.7479$, and $k = 3$. Therefore, $\mu' = -0.3095$. *Note* that when $\mu'$ is near zero or negative, the entry means would *not* be adjusted for incomplete block effects and that a randomized complete block analysis of covariance would be performed. *However*, in order to illustrate the computations, we proceed using the negative value for $\mu'$. For each of the nine entries, the computations are given in Table 5 for the adjusted totals. The adjusted totals are divided by the number of replicates, which is two for the simple lattice discussed here, to obtain the adjusted means. The sum of adjusted means is the total $Y_{\cdot\cdot\cdot}$ for the whole experiment.

The approximate average effective error variance for a difference between two adjusted means is (e.g., Federer, 1967, p. 323),

$$E_{yy}\left[1 + 2k\,\mu'\big/(k + 1)\right], \tag{9}$$

which for our example is

$$(1.0849)\left[1 + 2(3)(-0.3095)\big/4\right] = 1.0849\left[1 + 3(-0.3095)\big/2\right].$$

Table 5. Entry means adjusted for interblock information and covariates $X_1$ and $X_2$.

| Entry | Adjusted totals |
|---|---|
| 00 | $14-(-0.3095)\,[2-1.3986(0)-0.1884(-3)$ |
| | $-5-1.3986(-1)-0.1884(0)] = 14.3207$ |
| 10 | $6-(-0.3095)\,[3-1.3986(1)-0.1884(0)$ |
| | $-5-1.3986(-1)-0.1884(0)] = 6.6190$ |
| 20 | $7-(-0.3095)\,[3-1.3986(2)-0.1884(-2)$ |
| | $-5-1.3986(-1)-0.1884(0)] = 7.9352$ |
| 01 | $5-(-0.3095)\,[2-1.3986(0)-0.1884(-3)$ |
| | $-6-1.3986(-2)-0.1884(0)] = 5.1973$ |
| 11 | $10-(-0.3095)\,[3-1.3986(1)-0.1884(0)$ |
| | $-6-1.3986(-2)-0.1884(0)] = 10.4956$ |
| 21 | $5-(-0.3095)\,[3-1.3986(2)-0.1884(-2)$ |
| | $-6-1.3986(-2)-0.1884(0)] = 5.8119$ |
| 02 | $7-(-0.3095)\,[2-1.3986(0)-0.1884(-3)$ |
| | $+3-1.3986(0)-0.1884(5)] = 5.5691$ |
| 12 | $5-(-0.3095)\,[3-1.3986(1)-0.1884(0)$ |
| | $+3-1.3986(0)-0.1884(5)] = 3.8674$ |
| 22 | $13-(-0.3095)\,[3-1.3986(2)-0.1884(-2)$ |
| | $+3-1.3986(0)-0.1884(5)] = 12.1837$ |

Total        71.9999

However, since $\mu'$ is negative, a randomized complete error variance would be $\left\{\left(1-R^2_{B+E}\right)RV_{yy} = \right.$
$\left. 5.483626\right\} \Big/ (8-2) = 0.9139$. Then, the average variance of a difference between two means adjusted for regression would be $2(0.9139)/(r=2) = 0.9139$.

Note that if k is a prime or prime power, the entry means adjusted for interblock information may be computed directly rather than as described in equation (4). Formulas for doing this may be found in Federer (1967), Section IX–4.5 and Chapters XI and XII.

## 3. A Variance Component for Entry

It is desired to obtain an estimate of the component of variance for the 190 non-check entries in the experiment. Several procedures for doing this are possible. We shall use a sum of squares among adjusted entry means to estimate a component of variance associated with entry. Note that the effect of incomplete blocks and of covariates has been removed when obtaining the adjusted entry means. This means that the entry effect, the overall mean, and an error component are the only elements involved in an adjusted mean. Therefore, the sum of squares among the 190 non-check adjusted means $\bar{y}_{..gh}$ is

$$\sum_{gh=1}^{190} \bar{y}^2_{..gh} - \left(\sum_{gh} \bar{y}_{..gh}\right)^2 \Big/ 190 . \tag{10}$$

The expectation of this sum of squares is

$$E\left[\sum_{gh=1}^{190} \bar{y}^2_{..gh} - \left(\sum_{gh} \bar{y}_{..gh}\right)^2 \Big/ 190\right] = 189\left(\sigma^2_* + \sigma^2_T\right), \tag{11}$$

with 189 degrees of freedom. An estimate of $\sigma^2_*$ is one-half the average variance of a difference between two adjusted means, equation (9), i.e.,

$$\tfrac{1}{2} E'_{yy}\left\{1 + 2k\,\mu' \Big/ (k+1)\right\} = \hat{\sigma}^2_* \tag{12}$$

as the variance of a mean is one-half of the variance of a difference of two independent means. The above approximation ignored the correlation among adjusted means. Then, $\hat{\sigma}^2_T = $ the sum of squares in (11) divided by 189 minus (12).

## Example 2

Using the data in Example 1 and considering that entry 11 is check and that a variance component for the remaining 8 entries is desired, the sum of squares among the 8 entry-adjusted means from Table 5 is

$$\tfrac{1}{4}\Big\{14.3207^2 + 6.6190^2 + 7.9352^2 + 5.1973^2 + 5.8119^2 + 5.5691^2 + 3.8674^2$$

$$+ 12.1837^2\Big\} - \frac{(72 - 10.4956)^2}{4(8)} = 23.5545 ,$$

where $(1/2)^2 = 1/4$ was used because these are adjusted totals rather than means. Then,

$$\frac{23.5545}{8-1=7} - 0.9139/2 = 2.91 = \hat{\sigma}^2_T .$$

Since $\mu'$ was negative, the randomized complete error was used here. An error component of variance would be r $\sigma_*^2$, which would be 0.9139 for the example.

The next step is to consider all six sites and estimate variance components for sites by entry, entry over all sites, and an error component. Consider the two-way table of entry-adjusted means by sites. As pointed out in the Cochran and Cox (1957) reference above, there are various ways of combining data over sites. For our purposes, we use an unweighted means procedure where means from each site are given equal weights. An ANOVA table for the above two-way table of adjusted entry means by sites would be as given in Table 6. To obtain the expected values for the mean squares from the entry and entry X site lines in the ANOVA, it is assumed that the following linear model holds:

$$\overline{y}..._{ghi} = \mu^* + \gamma_i + \tau_{gh} + \gamma\tau_{ghi} + \epsilon^*_{ghi} \, , \tag{13}$$

where $\mu^*$ is a mean effect, $\gamma_i$ is an effect associated with site i, $\tau_{gh}$ is an average effect of entry gh over sites, $\gamma\tau_{ghi}$ is an interaction effect of entry gh with site i, and $\epsilon^*_{ghi}$ is a random error effect. For the maize experiment described above, consider that the $\tau_{gh}$ are identically and independently distributed (IID) with mean zero and variance $\sigma_\tau^2$, the $\gamma\tau_{ghi}$ are $\text{IID}\left(0, \ \sigma_{\gamma\tau}^2\right)$, the $\gamma_i$ are $\text{IID}\left(0, \sigma_\gamma^2\right)$, and the $\epsilon^*_{ghi}$ are $\text{IID}\left(0, \sigma_{\epsilon*}^2\right)$. Then the expectations of the means squares are given in Table 6. An estimate of $\sigma_{\epsilon*}^2$ would be obtained from the average of the estimated average variances of an adjusted mean over sites, i.e.,

$$\hat{\sigma}_{\epsilon*}^2 = \sum_{i=1}^{s} \hat{\sigma}_{*i}^2 \bigg/ s \, . \tag{14}$$

Estimation of the remaining variance components is then straightforward.

An alternate method would be to divide each of the adjusted means by their standard error of the mean before setting up the two-way table for entries and sites and obtaining the ANOVA in Table 6. Then, $\sigma_\epsilon^2$ may be taken as one and the $\hat{\sigma}_{\gamma\tau}^2$ and $\hat{\sigma}_\tau^2$ obtained from this table would be ratios of variance components.

Table 6. An analysis of variance (ANOVA) for adjusted entry means by sites, $\overline{y}_{..ghi}$.

| Source of variation | Degrees of freedom | Sum of squares | Expected value of mean square |
|---|---|---|---|
| Total | sv | $\sum\limits_{i=1}^{s} \sum\limits_{gh=1}^{v} \overline{y}^2_{.ghi}$ | — |
| Correction for mean | 1 | $\left( \sum\limits_{i} \sum\limits_{gh} \overline{y}_{.ghi} \right)^2 \Big/ sv = C$ | — |
| Site | (s–1) | $\sum\limits_{i} \left( \sum\limits_{gh} \overline{y}_{.ghi} \right)^2 \Big/ v - C = S$ | $\sigma^2_{\epsilon*} + \sigma^2_{\gamma\tau} + v\,\sigma^2_{\gamma}$ |
| Entry | (v–1) | $\sum\limits_{gh} \left( \Sigma\overline{y}_{.ghi} \right)^2 \Big/ s - C = V$ | $\sigma^2_{\epsilon*} + \sigma^2_{\gamma\tau} + s\,\sigma^2_{\tau}$ |
| Site × entry | (s–1)(v–1) | subtraction = I | $\sigma^2_{\epsilon*} + \sigma^2_{\gamma\tau}$ |

## 4. Genetic Advance

In any plant breeding program, it is desired to allocate resources in such a manner as to maximize progress in selecting for a characteristic such as yield. The resources are number of entries, number of replicates at a site, and number of sites. Formulas for doing this have been presented by Federer (1951) and Sprague and Federer (1951). The former paper does not consider costs; the latter paper does. Herein cost is ignored since the extra cost involved for CIMMYT would be the cost of sending separate seed packages and corresponding about the experiment at a given site. If these extra costs were available, the procedure in Sprague and Federer (1951) could be used.

One criterion to consider in allocating resources for an experiment is to consider that the total number of plots is to be a constant, say N. Then, the experimenter may vary r the number of replicates at a site, v the number of entries, and s the number of sites. Let

$$b = \frac{f_e - 2}{f_e} \left\{ \frac{\hat{\sigma}^2_{\gamma\tau}}{r\hat{\sigma}^2_{\epsilon*}} - \frac{2}{r(f_e - 2)} \right\} \tag{15}$$

and

$$d = \frac{f_e - 2}{f_e} \left\{ \frac{\hat{\sigma}^2_\tau}{r\hat{\sigma}^2_{\epsilon*}} \right\} \tag{16}$$

be the unbiased estimates of ratios of variance components (see Federer, 1951), where $f_e \doteq s(k^2 - 2k + 2)$. Then an estimate of the expected genetic advance is

$$\hat{G} = \frac{d\,\hat{\sigma}_\epsilon\,\overline{X}_v}{\sqrt{\frac{1}{s r} + \frac{b}{s} + d}}. \tag{17}$$

This is approximately equal to

$$\hat{G} = \frac{\hat{\sigma}^2_\tau\,\overline{X}_n}{\sqrt{\hat{\sigma}^2_\tau + \frac{\hat{\sigma}^2_{\gamma\tau}}{s} + \frac{\hat{\sigma}^2_\epsilon}{r s}}}, \tag{18}$$

where $\hat{\sigma}^2_\epsilon \left( = r\hat{\sigma}^2_{\epsilon*} \text{ from (14)} \right)$ is the average of the average effective error variances over sites. Since $f_e$ will be quite large $\left( 6(195-26) = 1014 \right)$, b and d are essentially ratios of variance components and hence (17) approaches (18). $\overline{X}_m$ is the expected value obtained by selecting the largest observation from a unit normal population. Expected values for the first, second, third, and fourth largest values from a

unit normal are given in Federer (1951) for values of $101 \leq v \leq 200$; the first five expected values are given for $41 \leq v \leq 100$. The largest value $\overline{\chi}_v$ for $v = 2, \cdots, 400$ and for $v = 500-1,000$ by tens may be found in Pearson and Hartley (1954, Table 27). Note that $\overline{\chi}_v$ is equal to one-half the range. We illustrate the above with a numerical example.

## Example 3

Genetic advances for various combinations of r, v, s, and ratios of variance components are presented in Table 7. Using formula (18), genetic advance is computed for the set $\sigma_T^2 = 10 = \sigma_\epsilon^2$ and $\sigma_{\gamma\tau}^2 = 5$, or ratios for b and d of $1/2$ and $1$. For the second set we use $\sigma_T^2 = 5$, $\sigma_{\gamma\tau}^2 = 2.5$, and $\sigma_\epsilon^2 = 10$, or $b = 1/4$ and $d = 1/2$. For the third set, $\sigma_T^2 = \sigma_{\gamma\tau}^2 = 1$ and $\sigma_\epsilon^2 = 10$ or $b = d = 1/10$. In Table 7, it may be noted that the maximum genetic advance for $G_1$, the first set, is attained when $v = 400$, $r = 1$, and $s = 6$. The maximum for $G_2$, the second set, is attained for $v = 200$, $r = 1$, and $s = 12$. Likewise, for set three, the maximum $G_3$ is attained for $v = 100$, $r = 1$, and $s = 24$. Note that selecting $N = 2400$ does not change the relative values; it was only selected as a common point for comparing all the allocations in Table 7.

As long as $\sigma_{\gamma\tau}^2 > 0$, larger G values will always be obtained for $r = 1$, i.e., a single replicate at each site. When costs are considered this usually will not be the case. Several allocations may give approximately the same values for G and of course, the cheapest and easiest would be selected. To illustrate, for $G_2$, $v = 200$, $r = 1$, and $s = 12$ gave almost the same value, 5.59, as $v = 400$, $r = 1$, and $s = 6$. It may be cheaper to obtain an additional 200 entries than to use an additional six sites; therefore, the latter allocation would be used. As another example for $v = 50$, note that $s = 12$ and $r = 4$ and $s = 24$ and $r = 2$ give almost the same G values as $s = 48$ and $r = 1$. The use of fewer sites would definitely be attractive.

Another point brought out by computations such as in Table 7, is that the smaller the value for d, the larger must be the values of r and s. For $d = 1$ v was 400, for $d = 1/2$ v was 200, and for $d = 1/10$ v was 100 in order to maximize G. It is necessary to have larger sample sizes, rs, to detect smaller differences among the entries. Selection in populations with large genetic variability and with a large mean is desirable in maximizing genetic progress.

Table 7. Genetic advance for various combinations of v, r, s, and values
of variance components for N = vrs = 2400.

| v | 4 | s | $\bar{X}_v$ | $G_1$ | $G_2$ | $G_3$ |
|---|---|---|---|---|---|---|
| 50 | 8 | 6 | 2.249 | 6.77 | 4.74 | 1.92 |
| | 4 | 12 | | 6.90 | 4.83 | 1.98 |
| | 2 | 24 | | 6.97 | 4.88 | 2.01 |
| | 1 | 48 | | 7.00 | 4.90 | 2.03 |
| 100 | 4 | 6 | 2.508 | 7.21 | 5.19 | 1.99 |
| | 2 | 12 | | 7.48 | 5.29 | 2.05 |
| | 1 | 24 | | 7.69 | 5.34 | 2.08* |
| 200 | 2 | 6 | 2.746 | 8.04 | 5.49 | 1.94 |
| | 1 | 12 | | 8.19 | 5.59* | 1.98 |
| 400 | 6 | 1 | 2.968 | 7.27 | 4.90 | 1.55 |
| | 3 | 2 | | 7.89 | 5.27 | 1.67 |
| | 2 | 3 | | 8.13 | 5.42 | 1.71 |
| | 1 | 6 | | 8.39* | 5.58 | 1.76 |
| 800 | 3 | 1 | 3.197 | 7.42 | 4.86 | 1.38 |
| | 1 | 3 | | 8.20 | 5.28 | 1.48 |

* Maximum value in column.

$\bar{X}_v$ value obtained as range/2 (see Pearson and Hartley, 1954, Table 27).

$G_1$: $\sigma_T^2 = 10$, $\hat{\sigma}_{T\gamma}^2 = 5$, $\hat{\sigma}_\epsilon^2 = 10$.

$G_2$: $\sigma_T^2 = 5$, $\sigma_{T\gamma}^2 = 2.5$, $\hat{\sigma}_\epsilon^2 = 10$.

$G_3$: $\sigma_T^2 = 1 = \sigma_{T\gamma}^2$, $\hat{\sigma}_\epsilon^2 = 10$.

## 5. Discussion

It may be necessary to adjust means for interblock regression as well as intrablock. If so, equation (1) would be changed to:

$$Y_{efg} = \mu + \rho_e + \pi_{ef} + \tau_g + \epsilon_{efg} + \sum_{d=1}^{p} \beta_d \left( X_{defg} - \bar{x}_{def\cdot} \right) + \sum_{d=1}^{p} \beta_d^* (\bar{x}_{def\cdot} - \bar{x}_{d\ldots}) , \tag{19}$$

where the $\beta_d^*$ are the incomplete block regression coefficients complete from the incomplete block (eliminating entry effects) sums of squares and cross products (see Table 1). Because of the relatively large number of degrees of freedom, 169, for intrablock error at each site, equation (1) was used for illustration and should correct for the covariates, but it will only be an approximation since the above equation may be more appropriate.

Since a large number of computations are involved, a computer software package such as SAS or GENSTAT should be used. As pointed out by Federer and Henderson (1979), Miles-McDermott et al. (1988), and Meredith et al. (1988), considerable caution must be used in using statistical packages when using covariance and different regressions such as intrablock and interblock. GENSTAT was the most successful and easiest to use. SAS should be able to handle the covariance analyses given in Section 2. In any event, any computer program selected should first be tried on the numerical example in Section 2.

A better treatment and experiment design than the one used would be to include the standard check variety once in each incomplete block. A rough guide on the number of times to replicate a check entry or variety when the comparison is the check variety against a new variety is the square root of the number of entries, i.e., $\sqrt{v}$. In this case, the check in each incomplete block would satisfy the above.

## 6. Literature Cited

Basford, K.E., C.E. McCulloch, and B. R. Murty (1992). Genotype by environment interactions. BU- -M in the Technical report Series of the Biometrics Unit, Cornell University, December.

Cochran, W.G. and G.M.Cox (1957). *Experimental Designs*, second edition. John Wiley & Sons, Inc., New York.

Federer, W.T. (1951). Evaluation of variance components from a group of experiments with multiple classifications. *Iowa Agricultural Expt. Station Research Bulletin* **380**: 241-310.

Federer, W.T. (1967). *Experimental Design — Theory and Application* (Indian edition). Oxford & IBH Publishing Co., Calcutta, Bombay, New Delhi. (Originally published by The Macmillan Co., 1955.)

Federer, W.T. and H.V. Henderson (1979). Covariance analyses of designed experiments × statistical packages: An update. In *Proceedings, Computer Science and Statistics: 12th Annual Symposium on the Interface* (J.F. Gentleman, editor). University of Waterloo, Waterloo, Ontario, pp. 228-235.

Federer, W.T. and M.P. Meredith (1992). Covariance analysis for split-plot and split-block designs. *The American Statistician* **46**: 155-162.

Federer, W.T. and T.P. Speed (1987). Measures of block design efficiency recovering interblock information. In *Transactions of the Fourth Army Conference on Applied Mathematics and Computing*, ARO Report 87-1, pp. 781-786.

Meredith, M.P., N.J. Miles-McDermott, and W.T. Federer (1988). The analysis of covariance for split-unit and repeated measures experiments using GLM procedure (II). In *Proceedings, SAS Users Group International, 13th Annual Conference*, pp. 1059-1064.

Miles-McDermott, N.J., W.T. Federer, M.P. Meredith, and Z.D. Feng (1988). The analysis of covariance for split-unit and repeated measures experiments using the GLM procedure in SAS/STAT© software (Part I). In *Proceedings, SAS Users Group International, 13th Annual Conference*, pp. 1053-1058.

Pearson, E.S. and H.O. Hartley (editors) (1954). *Biometrika Tables for Statisticians, Vol. I.* Published for the Biometrika Trustees at the University Press, Cambridge, England.

Rojas, B.A. and G.F. Sprague (1952). A comparison of variance components in corn yield trials: III. General and specific combining ability and their interaction with locations and years. *Agronomy J.* **44**: 462-466.

Snedecor, G.W. and W.G. Cochran (1980). *Statistical Methods,* 7th edition. The Iowa State University Press, Ames, Iowa.

Sprague, G.F. and W.T. Federer (1951). A comparison of variance components in corn yield trials: II. Error, year × variety, location × variety, and variety components. *Agronomy J.* **43**: 535-541.