Evaluation of variance components from a group of experiments with multiple classifications

Walter T. Federer
Iowa State College

May 1951

Follow this and additional works at: http://lib.dr.iastate.edu/researchbulletin

Part of the Agriculture Commons, and the Statistics and Probability Commons

Recommended Citation
Available at: http://lib.dr.iastate.edu/researchbulletin/vol30/iss380/1

This Article is brought to you for free and open access by the Iowa Agricultural and Home Economics Experiment Station Publications at Iowa State University Digital Repository. It has been accepted for inclusion in Research Bulletin (Iowa Agriculture and Home Economics Experiment Station) by an authorized editor of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Evaluation of Variance Components From A Group of Experiments With Multiple Classifications

By Walter T. Federer

Statistics Department

AGRICULTURAL EXPERIMENT STATION, IOWA STATE COLLEGE

Bureau of Agricultural Economics
United States Department of Agriculture
cooperating

RESEARCH BULLETIN 380 . . . . MAY, 1951 . . . . AMES, IOWA
CONTENTS

Summary .................................................................................................................. 245
Introduction ........................................................................................................ 247
Theoretical results .......................................................................................... 250

Combining groups of experiments ................................................................. 250
1. Unbiased estimates of the ratio of a variance component to the error mean square or variance .......... 250
2. Some characteristics of the averages of ratios of variance components ......................................................... 255

Average genetic advance due to the selection of the highest yielding individuals rather than a randomly selected set of individuals from a group of \( m \) strains ............................................ 258
3. Formulas for estimating the average genetic progress made by the selection of the highest yielding strain ....... 258
4. Formulas for estimating the average genetic progress made by selection of the \( k \) highest yielding strains ........ 260

Expectation of mean squares from a randomized complete block experiment containing all possible crosses of \( k \) lines...263
5. Expectation of mean squares .............................................................. 263
6. Variance for general combining ability .................................................. 267
7. Variance for specific combining ability .................................................. 268

Unbiased estimates of variance components from the weighted squares of means analysis for the two-way classification, with unequal numbers in the subclasses ......................................................... 269
8. Expectation of the mean squares for unequal numbers in each cell .............................................................. 269
9. Expectation of the mean squares for one or zero individuals in each cell ......................................................... 271

Other methods of analysis for the two-way classification with one or zero individuals in the subclasses .......... 273
10. The method of fitting constants ............................................................. 273
11. Missing plot formula and expectation of mean squares for \( k \) values missing in one group of the other classification ................................................................. 277
12. Missing plot formula and expectation of mean squares for \( k \) missing values for different entries in each of the two-way classifications ................................................................. 281
13. A simple method for obtaining the method of fitting constants analysis for a two-way classification with incomplete results ................................................................. 283
Experimental results and discussion .......................................................... 285
General description of the data .............................................................. 285
Experimental results on 302 corn experiments for yield .............. 287
  14. Analysis of the data .............................................................. 287
  15. Correlations among items in the analyses of variance ......... 288
  16. Frequency distribution of variance components and of the a's from 302 corn experiments .............................................. 288
  17. Average values of the a's for single crosses, top crosses and double crosses ................................................. 293

Data from experiments involving more than one tester .......... 295
  18. Frequency distribution of yields ................................................ 295
  19. Analyses of variance .............................................................. 295
  20. Estimation of the ratios of variance components ............ 300

Data from experiments involving more than one place or location .............................................................. 304
  21. Description of the data .............................................................. 304
  22. Relative value of combinations of strains, replicates and locations in corn yield trials .................................................. 305
  23. Relative value of combinations of replicates and locations in corn yield tests .................................................. 307

Literature cited .............................................................. 309
SUMMARY

The findings from the study on the development and application of methods for combining variance components from a group of experiments are:

1. Formulas for the expected value of the ratio of variance components from randomized block and split plot experiments were developed and applied to a set of yield data from experiments on corn. Variances for some of the ratios were developed and applied to the data.

2. The expectation of mean squares was found for two-way classifications of the following nature:
   (a) Randomized block experiments involving the $k(k-1)/2$ crosses among $k$ lines.
   (b) Randomized block experiments with disproportionate numbers in the subclasses by the weighted squares of means analysis.
   (c) Randomized block experiments with a number of missing or incomplete results by the method of fitting constants.
   (d) Randomized block experiments with the missing yields calculated and inserted in the table of yields. Two special cases were studied.

3. Although no generalized solution for missing plot values was found, the general solution was indicated in the original thesis (6). Formulas for calculating missing results are presented for two special cases.

4. The average ratio of the variety to the error variance components was highest for experiments involving single crosses, lowest for double crosses and intermediate for top crosses. These results follow expectation.

5. The unweighted mean of the unbiased estimates of a ratio of variance components was found to be satisfactory for estimating the population average ratio. When the number of experiments in the sample was greater than 30 to 40, the standard error of the unweighted mean was low enough for practical purposes. For the 11 tester experiments, the standard error of $\bar{b}$ (arithmetic mean of the unbiased ratios of error to line × tester variance components) was one-fourth of $\bar{b}$. This is larger than desired for most practical purposes, and either more homogeneous material or a larger number of experiments is required in order to reduce the variance of such an average ratio of variance components as $\bar{b}$. The ratio of the unweighted mean of variance components was more variable than the mean of the ratios.
6. For a fixed number of plots and for a simple randomized complete block design it was found that the greatest genetic advance in the selection of the highest yielding corn double crosses and single crosses was made when four and two replicates were used, respectively; the former type of corn hybrid exhibits relatively low and the latter relatively high genetic variation.

7. The present method for conducting the Iowa Corn Yield Tests is to use three locations or places with six replicates at a place. Considering the relative sizes of the error and the variety (single cross, top cross or double cross) by place variance components, it was found that six locations with two replicates at a place was 11 percent more efficient, and nine locations with one replicate at a place was as efficient as the present method. Also, one location with an infinite number of replicates was about one-half as efficient as the standard.

8. For a fixed number of plots the probability of obtaining the highest yielding corn hybrid was found to be higher if more locations and fewer replicates were used. Also six locations with two replicates at a place, with only two-thirds the number of plots, was found to be better for selecting high yielding hybrids than the present method. The formulas developed in the paper were used to obtain the above results.

Two methods were suggested for combining the variance components from the group of corn experiments. The methods are general and may be applied to similar data from various kinds of experiments.
Evaluation of Variance Components From
A Group of Experiments With
Multiple Classifications

By WALTER T. FEDERER

In conducting a replicated varietal yield trial, the experimenter usually is interested in obtaining the highest yielding variety or varieties from a group of varieties (or strains). A useful kind of information under such circumstances would be knowledge concerning the optimum allocation of replicates and varieties in any single experiment. Yates (23) and Perotti (12) have given a formula for determining the best combination of replicates and varieties in selecting the highest yielding variety or varieties from a simple randomized complete block design. For the more complex designs, formulas for determining the optimum allocation of varieties, replicates and locations (or years) would be useful.

The present paper deals with the summarization of the data from a group of experiments conducted as a part of the corn improvement program of the Iowa Agricultural Experiment Station and the Bureau of Plant Industry from 1940 through 1946 in Iowa. These data were made available through the kind permission of G. F. Sprague and J. L. Robinson. In the main, the objectives of this study are:

1. To summarize and evaluate the variance components from 302 corn yield trials.
2. To develop or extend the necessary statistical procedures for summarizing and evaluating these data.
3. To determine optimum relative values of the number of replicates and varieties and of replicates, varieties and locations for the selection of higher yielding single crosses, top crosses and double crosses.

The results are divided into two parts, the theoretical and

---

1Projects 890 and 163 of the Iowa Agricultural Experiment Station and the Bankhead-Jones Project, Bureau of Agricultural Economics, USDA. Condensed from a thesis submitted to the graduate faculty in partial fulfillment of requirements for the degree Doctor of Philosophy. The complete thesis, doctoral thesis No. 306, is on file in the library of Iowa State College. For their helpful suggestions and criticisms, the writer extends his thanks to W. G. Cochran and A. M. Mood, formerly of the Statistical Laboratory, Iowa State College, Ames, Iowa; to G. F. Sprague, senior agronomist, USDA, BPISAE, and collaborator, Iowa State College, Ames, Iowa; and to others who reviewed this impel. Dr. Sprague provided for a number of the calculations.

2Formerly collaborator, Iowa State College, in cooperation with the Bureau of Agricultural Economics, USDA.

3Senior agronomist, USDA, BPISAE, and collaborator with the Iowa Agricultural Experiment Station.

4Research professor, Iowa Agricultural Experiment Station.
the experimental. The theoretical results include the de-
velopment of statistical methods required for the summa-
ization of the data on variance components from corn ex-
periments, the solution of expectation of mean squares for
various analyses, and the development of formulas for vari-
ous combinations of missing plots, for various ratios of vari-
ance components, and for estimating average genetic ad-
vance. Two types of averages of ratios of variance compo-
nents were obtained and a third was investigated (6).

The experimental results include the application of the
statistical techniques to the corn data to determine the best
combination of strains and replicates or of strains, replicates
and locations in a single experiment which will, on the aver-
age, give the largest probability of obtaining the highest
yielding strains of corn, and to compare the averages of
ratios from the experiments.

For a given type of design, the error variances from
yield trials on corn hybrids would not, in general, be ex-
pected to be homogeneous since the fields on which they are
planted may have been under different managements. Also,
various climatic factors, such as cold weather and heavy
rains during germination, etc., have considerable effect upon
the experimental error for any particular experiment. The
data from the hybrid corn experiments studied in this paper
bear out these contentions. Likewise in an experiment in-
volving inbred lines of corn crossed with a group of unrelated
testers (or tested at a number of places), estimates of the
variance components for lines, \( \hat{\sigma}_v^2 \), and for testers by lines,
\( \hat{\sigma}_{vt}^2 \), may be expected to vary from experiment to experi-
ment; for example, a sample of lines from an open-pollinated
variety might be expected to be more variable than an equal
sized sample of lines from a high yielding F\(_2\) population. The
variation among lines might be increased considerably, due
to specific combining ability, by using an inbred line in place
of an open-pollinated variety as the tester parent. For any
set of corn strains which differ widely in maturity, the date
of planting or time of first killing frost may have consid-
erable effect on the estimate, \( \hat{\sigma}_v^2 \), of the variety variance com-
ponent.

For most yield trials grown over a large area, this de-
scription appears reasonable. Despite this variability, it
still is desirable to combine the data from previous years to
determine the most efficient procedure for future experi-
mentation.

An efficient method for summarizing the data on vari-
variance components from a large number of experiments of the nature described above was not immediately clear. Some methods could become extremely difficult due to the variation in number and kind of crosses and number of replicates and due to the fact that the distribution for a variance component is complex (13). There is, however, a fairly simple procedure for obtaining an unbiased estimate of the relative values of variance components from a group of experiments, if it is assumed that the experimenter is interested in the ratio of one of the variance components from the experiment to the error mean square or variance. In addition to the fact that it is possible to obtain an unbiased estimate of the ratio of variance components, the ratios can be combined in a simple manner if all the experiments have the same design and are assumed to give equal information regarding the relative sizes of the variance components. The average ratios of variance components are utilized for designing future experiments of the same nature as those summarized.

The ratios of variance components have an unknown distribution, which is a function of the $F$ distribution and the degrees of freedom, and it is desired to estimate the population parameter, which is also unknown. The individual ratios are random variables from the unknown distribution. In estimating a mean of the ratios, say $a_n$, where each ratio is an estimate of some parameter, say $a_n$, the mean, $\bar{a}$, will be subject to two sources of variation:

1. that due to the sampling errors in estimating the $a_n$, and
2. that due to the variation among the $a_n$.

If the variance among the $a_n$ is equal to or greater than the sampling errors associated with estimating the individual $a_n$, then the unweighted or arithmetic mean is more efficient than the weighted mean, in the sense that it has a lower variance. Cochran (2) has shown this to be the case for data which are not ratios but are simple estimates. He (2) further stated that in practice the situation was even more favorable for the arithmetic mean than indicated since the weights for a semi-weighted or weighted mean need to be estimated. This estimation results in a corresponding loss in information for these means with no corresponding loss in information on the unweighted mean.

The statistical procedure for combining the analyses from an experiment repeated at several times or locations has been given by Fisher (8) and Yates and Cochran (24). For this case, the analysis of variance must be applicable to each of the experiments, and the individual experiments must be
of equal precision. In the event that the individual experiments are not of equal precision, Smith (15), Cochran (2), and Yates and Cochran (24) have devised procedures for obtaining estimates of the treatment effects and the resulting tests of significance. Yates and Cochran (24) have given the method for combining the results of an experiment conducted at several places when the error degrees of freedom varied from place to place. Yates and Zacopanay (25) have considered a somewhat different situation, that is, the combination of the error variances from a group of cereal experiments in which the different experiments contained different cereals but any one experiment contained only varieties of a certain cereal; in their study the several experiments were combined to obtain an average estimate of the ratio of the sampling to the experimental error variances. The mean ratio of the sampling to the experimental error variances was used to determine the loss in information from sampling relative to complete harvesting of the individual plots in a yield trial.

Methods for estimating variance components have been described for particular types of designs by Yates and Zacopanay (25), Yates and Cochran (24), Winsor and Clarke (20), Satterthwaite (13, 14) and Crump (4, 5); the latter three authors dealt with the more general cases. Crump (5) gave variance components estimates for a two-way classification with disproportionate numbers in the subclasses for the method of "unweighted squares of means" and the method of "expected subclass numbers." Winsor and Clarke (20) and Crump (5) have worked out the expectations of the mean squares for an among-groups and within-groups analysis with disproportionate numbers in the groups. Another special case has been treated in a paper by Sprague and Tatum (17) where estimates of the variance in general and specific combining ability among the crosses within a group of $k$ lines were obtained for each of the lines involved.

THEORETICAL RESULTS

COMBINING GROUPS OF EXPERIMENTS

1. Unbiased estimates of the ratio of a variance component to the error mean square or variance.

The fact that the expected ratio is a function of Snedecor’s $F$ distribution is used in obtaining an unbiased estimate of the ratio of two variance components. For a randomized complete block experiment with $n$ replicates and $m$ crosses, the ratio of the variety variance component to the error
variance component may be expressed as

$$F = \frac{\sigma_{e}^2 + n\sigma_{v}^2}{\sigma_{e}^2 + n\sigma_{v}^2} = \frac{1 + na}{1 + na} \quad (1.1)$$

where \(a = \sigma_{r}^2 / \sigma_{e}^2\) and \(a = \sigma_{v}^2 / \sigma_{e}^2\). If \(f_e\) = degrees of freedom for error variance, \(f_v\) = degrees of freedom for the variety mean square, and \(E[\hat{a}] = a = \sigma_{v}^2 / \sigma_{e}^2\), then

$$E\left[ \frac{\sigma_{v}^2}{\sigma_{e}^2} \right] = \frac{f_v}{f_v - 2} + \frac{2}{n(f_e - 2)} \quad (1.2)$$

since the first moment of the \(F\) distribution is \(f_v/(f_v - 2)\) (19). Therefore, an unbiased estimate of \(a\) is

$$\hat{a} = \frac{f_v}{f_v - 2} \left\{ a - \frac{2}{n(f_e - 2)} \right\} \quad (1.3)$$

In the event that the expectation of the variety mean square is \(c\sigma_{e}^2 + n\sigma_{v}^2\), the unbiased estimate of the ratio of the variety to error variance components is

$$\hat{a} = \frac{f_v}{f_v - 2} \left\{ a - \frac{2c}{n(f_e - 2)} \right\} \quad (1.4)$$

where \(c\) is a constant. Also, if the variety mean square is of the form \(\sigma_{v}^2 + \frac{mn - k - n}{m - 1} \sigma_{e}^2\), then the unbiased estimate of the ratio of the variety to the error variance components is

$$\hat{a} = \frac{f_v}{f_v - 2} \left\{ a - \frac{2(m - 1)}{(mn - k - n)(f_e - 2)} \right\} \quad (1.5)$$

The ratio of the two variance components for experiments with multiple classifications may be obtained in a similar manner. For example, consider a randomized block experiment (design I, table 1) for which the testers and lines or varieties are randomized within the replicate. The yield equation for the \(i\)th plot of such an experiment will be assumed to have the form

$$y_{i_{gh}} = \mu + t_i + v_k + (rv)_{i_{g}} + (rt)_{i_{h}} + (vt)_{gh} + e_{i_{gh}} \quad (1.6)$$
### TABLE 1. ANALYSIS AND AVERAGE MEAN SQUARES FOR THREE DESIGNS.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Average values for mean squares</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Randomized Complete Block (I)</strong></td>
</tr>
<tr>
<td>Replicates</td>
<td>$n - 1$</td>
<td>$\sigma_e^2 + p\sigma_v^2 + m\sigma_r^2 + mnp\sigma_e^2$</td>
</tr>
<tr>
<td>Varieties</td>
<td>$(m - 1) = f_v$</td>
<td>$\sigma_e^2 + p\sigma_v^2 + m\sigma_v^2 + np\sigma_v^2$</td>
</tr>
<tr>
<td>Testers</td>
<td>$p - 1$</td>
<td>$\sigma_e^2 + m\sigma_r^2 + np\sigma_r^2 + mn\sigma_r^2$</td>
</tr>
<tr>
<td>Reps. x testers</td>
<td>$(n - 1)(p - 1)$</td>
<td>$\sigma_e^2 + m\sigma_r^2$</td>
</tr>
<tr>
<td>Var. x testers</td>
<td>$(m - 1)(p - 1) = f_{vt}$</td>
<td>$\sigma_e^2 + np\sigma_v^2$</td>
</tr>
<tr>
<td>Reps. x var.</td>
<td>$(n - 1)(m - 1) = f_{vr}$</td>
<td>$\sigma_e^2 + np\sigma_r^2$</td>
</tr>
<tr>
<td>Error</td>
<td>$(m - 1)(n - 1)(p - 1) = f_e$</td>
<td>$\sigma_e^2$</td>
</tr>
<tr>
<td>Total</td>
<td>$mnmp - 1$</td>
<td>Total variance</td>
</tr>
</tbody>
</table>

**Split-Plot Design (II)**

| Replicates          | $(n - 1)$          | $\sigma_e^2 + m\sigma_r^2 + np\sigma_e^2$ |
| Testers             | $(p - 1)$          | $\sigma_e^2 + m\sigma_r^2 + np\sigma_r^2 + mn\sigma_r^2$ |
| Error (a)           | $(n - 1)(p - 1)$   | $\sigma_e^2 + m\sigma_r^2$ |
| Lines (or var.)     | $(m - 1) = f_v$    | $\sigma_e^2 + np\sigma_v^2$ |
| Lines x testers     | $(m - 1)(p - 1) = f_{vt}$ | $\sigma_e^2 + np\sigma_v^2$ |
| Error (b)           | $p(m - 1)(n - 1) = f_e$ | $\sigma_e^2$ |
| Total               | $mnmp - 1$         | Total variance |

**Split-Plot Design (III)**

| Places | $p - 1$ | $\sigma_e^2 + m\sigma_r^2 + np\sigma_r^2 + mn\sigma_r^2$ |
| Reps. within a place | $p(n - 1)$ | $\sigma_e^2 + m\sigma_r^2$ |
| Varieties | $(m - 1) = f_v$ | $\sigma_e^2 + np\sigma_v^2 + np\sigma_r^2$ |
| Variety x place | $(m - 1)(p - 1) = f_{1v}$ | $\sigma_e^2 + np\sigma_v^2$ |
| Error | $p(m - 1)(n - 1) = f_e$ | $\sigma_e^2$ |
| Total | $mnmp - 1$ | Total variance |
where $\mu$ is the general mean, $r_i$ the effect common to the $i$th replicate, $t_h$ the effect common to the $h$th tester, $v_g$ the effect common to the $g$th line, $(rv)_{ig}$ the effect common to the $g$th line in the $i$th replicate, $(rt)_{ih}$ the effect common to the $h$th tester in the $i$th replicate, $(vt)_{gh}$ the effect common to the $g$th line crossed with the $h$th tester, and $e_{igh}$ is an effect common to all plots in the experiment; $i = 1, 2, \ldots, n$; $g = 1, 2, \ldots, m$; $h = 1, 2, \ldots, p$.

For a split-plot design composed of $n$ replicates each containing $p$ testers as the whole plot and $m$ varieties or lines crossed with each tester as the split or subplot (design II, table 1) the individual plot yield will be assumed to be expressed by the linear equation

$$y_{igh} = \mu + r_i + t_h + (rt)_{ih} + v_g + (vt)_{gh} + e_{igh}. \quad (1.7)$$

Likewise for a split-plot design consisting of $m$ varieties or lines in each of $n$ replicates, planted at $p$ locations (design III, table 1), the yield of the individual plot will be assumed to have the form

$$y_{igh} = \mu + l_i + (lr)_{ih} + v_g + (lv)_{gh} + e_{igh}. \quad (1.8)$$

where $i = (h-1)n + 1$, $(h-1)n + 2$, $\ldots$, $hn$; $g = 1, 2, \ldots, m$; $h = 1, 2, \ldots, p$; and the individual regression coefficients are designated in a similar manner as that for design I. The average values for the various mean squares are given in table 1. The expectations of mean squares for other complex designs may be obtained in a similar manner. The above analyses will be sufficient for the purposes of this study.

For simplicity let $f_e$ be the degrees of freedom for the error variance and $f_{vt}$ the degrees of freedom for the variety $\times$ tester (or place) mean square. For designs I and II (table 1) the unbiased estimate of the ratio of the line or variety by tester variance component to the error variance component is

$$\hat{b} = \frac{f_e - 2}{f_e} \left\{ \frac{\hat{\sigma}_{vt}^2}{\hat{\sigma}_e^2} - \frac{2}{n(f_e - 2)} \right\}$$

$$= \frac{f_e - 2}{f_e} \left\{ b - \frac{2}{n(f_e - 2)} \right\}. \quad (1.9)$$

A similar formula is obtained for the unbiased estimate of the ratio $\beta = \frac{\sigma_{vr}^2}{\sigma_v^2}$ from design III. Likewise, an unbiased
estimate of \( \gamma = \frac{\sigma_{v}^2}{\sigma_{e}^2} \) (design I) is

\[
\hat{c} = \frac{f_e - 2}{t_e} \left\{ \frac{\hat{\sigma}_{v}^2}{\hat{\sigma}_{e}^2} - \frac{2}{p(f_e - 2)} \right\}
\]

\[
= \frac{f_e - 2}{t_e} \left\{ c - \frac{2}{p(f_e - 2)} \right\} .
\]

(1.10)

The unbiased estimate of \( \frac{\sigma_{v}^2}{\sigma_{e}^2} \) from designs II or III is

\[
\hat{d} = \frac{f_e - 2}{t_e} \left\{ \frac{\hat{\sigma}_{v}^2}{\hat{\sigma}_{e}^2} \right\} = \frac{d(f_e - 2)}{t_e} .
\]

(1.11)

For design I, the unbiased estimate of the ratio of the variety variance component to the error variance component is

\[
\hat{e} = \frac{f_e - 2}{t_e} \left\{ \frac{\hat{\sigma}_{v}^2}{\hat{\sigma}_{e}^2} + \frac{2}{np(f_e - 2)} \right\}
\]

\[
= \frac{f_e - 2}{t_e} \left\{ e + \frac{2}{np(f_e - 2)} \right\} .
\]

(1.12)

The unbiased estimate of the ratio \( \frac{\sigma_{v}^2}{\sigma_{e}^2 + n\sigma_{v1}^2} \) for designs I and II is

\[
\hat{f} = \frac{f_{vt} - 2}{t_{vt}} \left\{ \frac{\hat{\sigma}_{v}^2}{\hat{\sigma}_{e}^2 + \hat{n}\sigma_{v1}^2} - \frac{2}{np(f_{vt} - 2)} \right\}
\]

\[
= \frac{f_{vt} - 2}{t_{vt}} \left\{ f - \frac{2}{np(f_{vt} - 2)} \right\} .
\]

(1.13)

A similar formula is obtained for the ratio \( \frac{\sigma_{v}^2}{\sigma_{e}^2 + n\sigma_{v1}^2} \) in design III.

The ratios of other variance components to the error variance component are obtained in a similar manner.

A simple method for combining the ratios of variance components from a group of experiments is to obtain an arithmetic mean of the ratios, but this mean would be biased. Therefore, in order to obtain an unbiased mean of the ratios,
say \( a_u \), use the unbiased estimates of the ratios, \( \hat{a}_u \), and obtain an arithmetic mean of these, thus

\[
\bar{a} = \frac{1}{N} \sum_{u=1}^{N} \hat{a}_u.
\]  

(1.14)

If

\[
\bar{a}' = \frac{1}{N} \sum_{u=1}^{N} a_u,
\]  

(1.15)

then \( \bar{a} \) and \( \bar{a}' \) will not approach the same mean \( a \) since \( \bar{a}' \) is biased. Furthermore, the variance of \( a_u \) is larger than the variance of \( \hat{a}_u \) by a factor \( \frac{f_e^2}{(f_e - 2)^2} \) [see formula (2.2)] which becomes important for \( f_e \) small.

The maximum likelihood estimate of a ratio, say \( \hat{a}_u \), of two variance components is the ratio of the two components (6). In order to combine these estimates it is necessary to assume that the \( \hat{a}_u \) are all estimates of the same parameter, say \( a \). This assumption does not hold for the corn data and hence this method is not included in the present paper. For a discussion of maximum likelihood solutions of ratios of variance components, the reader is referred to the original thesis (6).

2. Some characteristics of the averages of ratios of variance components.

For \( a = \frac{\hat{c}_{\sigma_e^2} + \hat{n_{\sigma_v^2}}}{\hat{c}_{\sigma_e^2}} - \frac{c}{n} \), where the error, \( \sigma_e^2 \), and the variety, \( c_{\sigma_e^2} + n_{\sigma_v^2} \), mean squares have \( f_e \) and \( f_v \) degrees of freedom, respectively, the variance of \( a \) is

\[
V(a) = E[a^2] - [Ea]^2 =
\]

\[
\frac{2f_e^2}{f_v(f_v - 4)(f_e - 2)^2} \left\{ \hat{a} + \frac{c}{n} \right\}^2.
\]  

(2.1)

The variance of \( \hat{a} \) is

\[
V(\hat{a}) = \frac{(f_e - 2)^2}{f_e^2} V(a)
\]

\[
= \frac{2(f_e + f_v - 2)}{f_v(f_v - 4)f_e^2} \left\{ \frac{\hat{c}_{\sigma_e^2} + \hat{n_{\sigma_v^2}}}{\hat{c}_{\sigma_e^2}} \right\}^2.
\]  

(2.2)
Making use of the above results the variance of \( \bar{a} \) is (assuming all \( a_u = a \))

\[
V(\bar{a}) = \frac{1}{N^2} \sum_{u=1}^{N} V(\hat{a}_u) =
\]

\[
\frac{1}{N^2} \sum_{u=1}^{N} \left[ \frac{2(f_u + f_v - 2)(f_u - 2)^2}{f_v(f_v - 4)f_v^2} \right] \left[ \frac{\hat{\sigma}_u^2 + n\sigma_v^2}{n\hat{\sigma}_v^2} \right]^2
\]

(2.3)

where \( \bar{a} = \frac{1}{N} \sum_{u=1}^{N} \hat{a}_u \) and the \( \hat{a}_u \) are independent (19).

Since the maximum likelihood solution gives an efficient estimate of the parameter \( a \), any other estimate may be compared with it to determine its relative efficiency. To determine the efficiency of the unbiased estimate \( \bar{a} \) (1.14) as compared to \( \hat{a} \), the maximum likelihood estimate (6), the variances of the two estimates are compared, thus

\[
\frac{V(\bar{a})}{V(\hat{a})}
\]

above variances are computed under the assumption that the \( a_u = a \). This assumption does not hold for the data considered in this paper. When the \( a_u \neq a \) and where the variance among the \( a_u \) is as large or larger than the sampling errors of the individual \( a_u \), the unweighted mean is preferable to the weighted mean. The variance of the unweighted mean, formula (1.14), may be approximated by the formula

\[
\frac{N}{u=1} \frac{(\hat{a}_u - \bar{a})^2}{N(N-1)}.
\]

These results have been obtained by Cochran (2) for simple estimates of the mean effect.

As a third average, the individual components may be averaged and the average ratio obtained as the ratio of the averages. Thus, for

\[
\bar{\sigma}_v^2 = \frac{1}{N} \sum_{u=1}^{N} (\hat{\sigma}_v^2)_u \quad \text{and} \quad \bar{\sigma}_v^2 = \frac{1}{N} \sum_{u=1}^{N} (\hat{\sigma}_v^2)_u,
\]

\[
\bar{a}'' = \frac{\bar{\sigma}_v^2}{\bar{\sigma}_v^2}.
\]

(2.4)

This average will have a variance which is approximated by
\[ V(\bar{a}^u) = \left\{ \frac{\sigma^2}{\bar{a}^u} \right\}^2 \left\{ \frac{V(\hat{\sigma}^2)}{\bar{a}^u} + \frac{V(\hat{\sigma}_{e}^2)}{\bar{a}^u} - \frac{2 \text{ cov } \sigma^2 \bar{a}^u}{N \bar{a}^u} \right\} \] (2.5)

where

\[ V(\hat{\sigma}^2) = \frac{1}{N^2} \sum_{u=1}^{N} V(\hat{\sigma}_{e}^2) \] (2.6)

and

\[ V(\hat{\sigma}_{e}^2) = \frac{1}{N^2} \sum_{u=1}^{N} V(\hat{\sigma}_{e}^2) \, u. \] (2.7)

The individual variances \( V(\hat{\sigma}_{e}^2) \) and \( V(\hat{\sigma}_{v}^2) \) have been given by Crump (4) as

\[ V(\hat{\sigma}_{e}^2) = \frac{2}{n^2} \left\{ \frac{(\hat{\sigma}_{e}^2 + n\hat{\sigma}_{e}^2)^2}{f_e + 2} + \frac{\hat{\sigma}_{e}^2}{f_e + 2} \right\} \] (2.8)

and

\[ V(\hat{\sigma}_{v}^2) = \frac{2\hat{\sigma}_{e}^2}{f_e} \] (2.9)

where the symbols are as defined above. An approximate variance may be obtained by considering the individual variance components as normally distributed independent variates and then proceeding in the same manner as for normal variate theory.

To test the hypothesis that the individual ratios \( \hat{a}_u \) may have come from the same population, the following test of significance is proposed:

\[ \chi^2 (N - 1 \text{ d.f.}) = \sum_{u=1}^{N} \frac{(\hat{a}_u - \bar{a})^2}{NV(\bar{a})} \] (2.10)

The above test is an approximation since \( \hat{a}_u \) are assumed to be normally distributed or nearly so. If the \( \hat{a}_u \) are not normally distributed, then some transformation, such as log

\[ \left\{ \frac{n + 1}{n} + \hat{a}_u \right\}, \] may be used to make them so.
3. Formulas for estimating the average genetic progress made by the selection of the highest yielding strain.

Since such characters as yield, disease resistance, etc. usually are controlled by a large number of factors, it is often necessary to test large numbers of strains in an experiment. The greater the number of strains tested, the larger will be the probability for obtaining a particularly favorable combination of a large number of factors. Also, for a fixed number of experimental plots, the greater the number of strains tested the less accurate will be the comparisons among the strains. An increase in the block size is usually accompanied by an increase in the error variance. Therefore, if the increase in the error variance is neglected and if it is assumed that the number of plots is fixed, then the following procedure for obtaining the formula given by Perotti (12) may be used to estimate the best combination of strains and replicates for a given ratio of the error and variety variance components.

For m strains or varieties and n replicates, the analysis of variance on an experimental mean basis is:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Average value of mean squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varieties</td>
<td>m-1</td>
<td>( \frac{\sigma_e^2}{mn} + \frac{\sigma_r^2}{m} )</td>
</tr>
<tr>
<td>Error</td>
<td>(n-1)(m-1)</td>
<td>( \frac{\sigma_e^2}{mn} )</td>
</tr>
</tbody>
</table>

The intraclass correlation or the regression of phenotype on genotype is \( \frac{\sigma_e^2}{\sigma_e^2 + \frac{\sigma_v^2}{n}} \), the standard deviation is \( \sigma_v + \frac{\sigma_e^2}{n} \) and the average value for selecting the highest yielding variety instead of a randomly selected one from a sample of m from a normal population with mean zero and unit variance is \( \bar{x}_m \). Now \( \bar{x}_m \sqrt{\sigma_v^2 + \frac{\sigma_e^2}{n}} \) would be the average gain in standard deviation units due to the selection of the apparently best individual. Since the correlation between phenotype and
genotype, \( \frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_e^2}{n}} \), is not perfect, then the average gain must be weighed by the correlation between observed and actual, thus

\[
\frac{\sigma_v^2 \bar{X}_m}{\sigma_v^2 + \frac{\sigma_e^2}{n}} \sqrt{\sigma_v^2 + \frac{\sigma_e^2}{n}} = \frac{\sigma_v^2 \bar{X}_m}{\sqrt{\sigma_v^2 + \frac{\sigma_e^2}{n}}} = \sqrt{\alpha + \frac{1}{n}}
\]

(3.1)

The above procedure may be extended to cover more complex situations. The analysis for a split plot design (table 1), on an experimental mean basis, is

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Average value of mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varieties ((m - 1))</td>
<td>(\sigma_v^2/pmn + \sigma_{vt}^2/pm + \sigma_e^2/m)</td>
<td></td>
</tr>
<tr>
<td>Varieties x testers ((p - 1)(m - 1))</td>
<td>(\sigma_v^2/pmn + \sigma_{vt}^2/pm)</td>
<td></td>
</tr>
<tr>
<td>Error ((p(m - 1)(n - 1))</td>
<td>(\sigma_v^2/pmn)</td>
<td></td>
</tr>
</tbody>
</table>

In place of testers, years or locations may be used in a similar analysis. The intraclass correlation is \(\frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_{vt}^2}{p} + \frac{\sigma_e^2}{np}}\) and the standard deviation is \(\sqrt{\sigma_v^2 + \frac{\sigma_{vt}^2}{p} + \frac{\sigma_e^2}{np}}\). Thus the average gain or genetic advance due to the selection of the highest yielding instead of a randomly selected variety from a sample of \(m\) lines is

\[
\frac{\sigma_v^2 \bar{X}_m}{\sqrt{\sigma_v^2 + \frac{\sigma_{vt}^2}{p} + \frac{\sigma_e^2}{np}}} = \delta \frac{\sigma_v \bar{X}_m}{\sqrt{\delta + \frac{\beta}{p} + \frac{1}{np}}}.
\]

(3.2)

Likewise for a design in which a group of \(m\) lines are crossed with \(p\) testers and the resulting crosses grown in \(n\) completely randomized blocks, the following pertinent part of the analysis (table 1) is
In this case the average progress made by selecting the highest yielding variety is

\[ \frac{\sigma^2_{x\text{m}}}{\sqrt{\sigma^2 + \frac{\sigma^2_{rt}}{n} + \frac{\sigma^2_{vt}}{p} + \frac{\sigma^2_e}{np}}} \]

which reduces to the preceding formula if \( \sigma^2_{rv} = 0 \).

The above process may be extended to other situations by the method outlined above.

4. **Formulas for estimating the average genetic progress made by selection of the \( k \) highest yielding strains.**

In section 3, formulas were given for obtaining the average genetic advance due to the selection of the highest yielding rather than a randomly selected line from a sample of \( m \) lines. The expected value for the largest deviate may be obtained from the table of ranges given by Tippett (18). These tables were reproduced by Karl Pearson (11) in table XXII and in part by Snedecor (16). If it is desired to obtain the expected value for the \( k \) highest yielding strains from a sample of size \( m \), the expected values for the highest, the next highest, etc., down to the \( k \)th highest yielding individuals are averaged. For example, the expected value for the two highest yielding strains from a sample of size 50 is

\[ \frac{(2.25 + 1.85)\sigma}{2} = 2.05\sigma \]

where \( \sigma \) is the standard deviation and the values 2.25 and 1.85 were obtained from table XX of Fisher and Yates (9). Unfortunately the expected values for the second largest, the third largest, etc., have not been tabulated for samples of \( m \) greater than 50.
Since the formula for the expected value of the kth largest deviate (9, 19) involved tedious numerical integration, some simpler solution is desired. The values for the expected value of the range have been tabled by Tippett (18) for values of m ranging from 2 to 1,000. The expected value of the largest deviate is one-half of the range. These values are given in table 2 for m = 41 to 100 and in table 3 for m = 101 to 200.

The expected value of the largest deviate from a sample of size m is given as (9, 19),

\[
E(x_{m,1}) = m \int_{\infty}^{\infty} \int_{\infty}^{x_1} f(x) \, dx \left( \int_{\infty}^{x_1} f(x) \, dx \right)^{m-1} f(x_1) \, dx_1
\]

\[= m \int x F^{m-1} \, dF. \quad (4.1)\]

Formula (4.1) may be written as

\[
E(x_{m,1}) = m \int x F^{m-2} \, dF - m \int x F^{m-2} (1-F) \, dF
\]

\[= \frac{m}{m-1} E(x_{m-1,1}) - \frac{1}{m-1} E(x_{m,2}). \quad (4.2)\]

Therefore, the expected value for the second largest member from a sample of size m may be obtained from the largest members of samples of size m and m - 1, thus

\[
E(x_{m,2}) = mE(x_{m-1,1}) - (m - 1)E(x_{m,1}). \quad (4.3)\]

The recursion formulas for obtaining the third, fourth and fifth largest deviates, respectively, were found by the same device to be

\[
E(x_{m,3}) = \frac{1}{2} \left\{ mE(x_{m-1,2}) - (m - 2)E(x_{m,2}) \right\}, \quad (4.4)\]

\[
E(x_{m,4}) = \frac{1}{3} \left\{ mE(x_{m-1,3}) - (m - 3)E(x_{m,3}) \right\}, \quad (4.5)\]

\[
E(x_{m,5}) = \frac{1}{4} \left\{ mE(x_{m-1,4}) - (m - 4)E(x_{m,4}) \right\}. \quad (4.6)\]

Recursion formulas for other ordered deviates may be set up in a similar manner. However, due to the fact that the range tabulated by Tippett (18) was correct to four decimal places only, it is not possible to obtain the expected values of the third largest deviates for samples of 100 or less correct to more than one decimal place. If the expected value
<table>
<thead>
<tr>
<th>Sample size</th>
<th>Value of deviate</th>
<th>Sample size</th>
<th>Value of deviate</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>2.171 1.765 1.530 1.36 1.22</td>
<td>71</td>
<td>2.383 2.007 1.793 1.63 1.52</td>
</tr>
<tr>
<td>42</td>
<td>2.180 1.776 1.542 1.37 1.23</td>
<td>72</td>
<td>2.388 2.013 1.799 1.64 1.53</td>
</tr>
<tr>
<td>43</td>
<td>2.190 1.787 1.554 1.38 1.25</td>
<td>73</td>
<td>2.393 2.019 1.806 1.65 1.57</td>
</tr>
<tr>
<td>44</td>
<td>2.199 1.797 1.567 1.40 1.26</td>
<td>74</td>
<td>2.398 2.025 1.813 1.66 1.59</td>
</tr>
<tr>
<td>45</td>
<td>2.208 1.807 1.580 1.41 1.27</td>
<td>75</td>
<td>2.403 2.031 1.819 1.66 1.54</td>
</tr>
<tr>
<td>46</td>
<td>2.216 1.817 1.593 1.42 1.28</td>
<td>76</td>
<td>2.408 2.036 1.825 1.66 1.54</td>
</tr>
<tr>
<td>47</td>
<td>2.225 1.827 1.603 1.43 1.30</td>
<td>77</td>
<td>2.413 2.042 1.832 1.67 1.54</td>
</tr>
<tr>
<td>48</td>
<td>2.233 1.837 1.614 1.44 1.31</td>
<td>78</td>
<td>2.417 2.047 1.838 1.68 1.55</td>
</tr>
<tr>
<td>49</td>
<td>2.241 1.847 1.624 1.45 1.32</td>
<td>79</td>
<td>2.422 2.052 1.844 1.69 1.55</td>
</tr>
<tr>
<td>50</td>
<td>2.249 1.856 1.634 1.46 1.33</td>
<td>80</td>
<td>2.427 2.057 1.850 1.69 1.56</td>
</tr>
<tr>
<td>51</td>
<td>2.257 1.864 1.644 1.48 1.34</td>
<td>81</td>
<td>2.431 2.063 1.855 1.70 1.56</td>
</tr>
<tr>
<td>52</td>
<td>2.264 1.873 1.654 1.49 1.36</td>
<td>82</td>
<td>2.436 2.068 1.861 1.71 1.56</td>
</tr>
<tr>
<td>53</td>
<td>2.272 1.881 1.662 1.50 1.37</td>
<td>83</td>
<td>2.440 2.073 1.867 1.71 1.57</td>
</tr>
<tr>
<td>54</td>
<td>2.279 1.889 1.669 1.51 1.38</td>
<td>84</td>
<td>2.445 2.077 1.872 1.72 1.57</td>
</tr>
<tr>
<td>55</td>
<td>2.286 1.897 1.677 1.52 1.39</td>
<td>85</td>
<td>2.449 2.082 1.877 1.73 1.58</td>
</tr>
<tr>
<td>56</td>
<td>2.293 1.905 1.685 1.53 1.40</td>
<td>86</td>
<td>2.453 2.087 1.883 1.73 1.58</td>
</tr>
<tr>
<td>57</td>
<td>2.300 1.912 1.692 1.54 1.42</td>
<td>87</td>
<td>2.457 2.092 1.888 1.74 1.59</td>
</tr>
<tr>
<td>58</td>
<td>2.306 1.920 1.700 1.55 1.43</td>
<td>88</td>
<td>2.462 2.096 1.893 1.75 1.60</td>
</tr>
<tr>
<td>59</td>
<td>2.313 1.927 1.708 1.56 1.44</td>
<td>89</td>
<td>2.466 2.101 1.898 1.75 1.61</td>
</tr>
<tr>
<td>60</td>
<td>2.319 1.935 1.715 1.56 1.45</td>
<td>90</td>
<td>2.470 2.105 1.902 1.76 1.61</td>
</tr>
<tr>
<td>61</td>
<td>2.326 1.942 1.722 1.57 1.46</td>
<td>91</td>
<td>2.474 2.110 1.907 1.77 1.62</td>
</tr>
<tr>
<td>62</td>
<td>2.332 1.949 1.729 1.57 1.47</td>
<td>92</td>
<td>2.478 2.114 1.912 1.77 1.63</td>
</tr>
<tr>
<td>63</td>
<td>2.338 1.956 1.737 1.58 1.48</td>
<td>93</td>
<td>2.482 2.119 1.916 1.78 1.64</td>
</tr>
<tr>
<td>64</td>
<td>2.344 1.963 1.744 1.59 1.49</td>
<td>94</td>
<td>2.485 2.123 1.920 1.78 1.65</td>
</tr>
<tr>
<td>65</td>
<td>2.350 1.969 1.751 1.60 1.49</td>
<td>95</td>
<td>2.489 2.127 1.925 1.79 1.66</td>
</tr>
<tr>
<td>66</td>
<td>2.355 1.976 1.758 1.60 1.50</td>
<td>96</td>
<td>2.493 2.131 1.929 1.80 1.67</td>
</tr>
<tr>
<td>67</td>
<td>2.361 1.983 1.765 1.61 1.51</td>
<td>97</td>
<td>2.497 2.136 1.933 1.80 1.68</td>
</tr>
<tr>
<td>68</td>
<td>2.367 1.989 1.772 1.61 1.51</td>
<td>98</td>
<td>2.500 2.140 1.937 1.81 1.69</td>
</tr>
<tr>
<td>69</td>
<td>2.372 1.995 1.779 1.62 1.52</td>
<td>99</td>
<td>2.504 2.144 1.941 1.81 1.70</td>
</tr>
<tr>
<td>70</td>
<td>2.377 2.001 1.786 1.63 1.52</td>
<td>100</td>
<td>2.508 2.148 1.945 1.81 1.70</td>
</tr>
</tbody>
</table>
of every fifth or sixth largest deviate were obtained correctly to several significant figures, say 8 to 10, by numerical integration, then it would be possible to use the above recursion formula and obtain the values for the ordered deviates in between by successive approximations.

Because data on expected values of deviates are not available to several significant figures, the recursion formulas are not satisfactory in themselves. This fact is brought out by a study of the values given in tables 3 and 4 of the thesis (6). However, by computing a fifth degree regression on the values given by Tippett (18) for the largest deviate, equal one-half the expected range, the values were computed to nine significant figures. Recursion formula (4.3) was then applied to these values and the expected values for the second largest deviates obtained. A fifth degree regression was fitted to the latter values to obtain the second largest deviates correct to seven decimals. Recursion formula (4.4) was applied to obtain the third largest deviates. This process was continued for the fourth and fifth largest deviates. These values are given in tables 2 and 3 to four significant figures for the first, second and third largest deviates and to three significant figures for the fourth and fifth.

These data are suitable for most practical conditions encountered. The agreement of the values in tables 2 and 3 with those given by Fisher and Yates (9) was good, but it is suggested that the accuracy of this method be tested further before many more additional expected values are computed.

A rough approximation of the expected value of the largest \( k \) individuals (\( k \) less than \( m/2 \)) may be obtained from the next to the last column of Pearson's (11) table II. This approximation will approach the true value as the sample size \( m \) increases. If \( k > m/2 \), then Pearson's (10) table XI should be used.

For the value \( \bar{x}_m \) in formulas (3.1), (3.2) and (3.3), an average of the values in table 2 may be used to obtain the average genetic progress made by selecting the highest two, three, four or five strains from samples of size 41 to 100. Likewise, the values in table 3 may be used to obtain the average advance due to the selection of the two, three or four highest yielding strains from samples of size 101 to 200.

**EXPECTATION OF MEAN SQUARES FROM A RANDOMIZED COMPLETE BLOCK EXPERIMENT CONTAINING ALL POSSIBLE CROSSES OF \( k \) LINES**

5. **Expectation of mean squares.**

A number of breeding experiments are conducted using all possible crosses among a group of \( k \) lines. In some cases
the reciprocal crosses are identical and hence are not included in the experiment, for example, yield trials involving single-cross or double-cross corn hybrids. Such an experiment would have \( k(k-1)/2 \) crosses, and if the design were a randomized complete block, the yield of an individual
plot will be assumed to be represented by the linear equation

\[ y_{ij} = \mu + r_i + v_j + e_{ij}, \quad (5.1) \]

where \( i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, k \); \( k(k - 1)/2 \); \( \mu \) is the general mean, \( r_i \) is an effect common to the \( i \)th replicate, \( v_j \) is an effect common to the \( j \)th variety, and \( e_{ij} \) is an effect common to all plots. If it is desired to consider the general or over-all effect of \( k \) lines, then the yield of the \( i \)th plot is expressed as

\[ y_{i\text{gh}} = \mu + r_i + l_g + l_h + (lt)_{gh} + e_{i\text{gh}}, \quad (5.2) \]

where \( g \neq h = 1, 2, \ldots, k \); the double subscript \( gh \) runs from \( g < h = 2, 3, \ldots, k \); \( l_g \) is the effect common to the \( g \)th line, \( l_h \) is the effect common to the \( h \)th line, \( (lt)_{gh} \) is the effect common to the cross of the \( g \)th by the \( h \)th line, and \( e_{i\text{gh}} \) is an effect common to all plots.

The analysis of variance, considering the line effects, for such an experiment is:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicates</td>
<td>( n - 1 )</td>
<td>( \sum_{i=1}^{n} \frac{2Y_i^2}{k(k-1)} - \frac{2Y^2}{nk(k-1)} )</td>
</tr>
<tr>
<td>Varieties</td>
<td>( (m - 1) = \frac{k(k-1)}{2} - 1 )</td>
<td>( \sum_{g&lt;h=2}^{k} \frac{Y_{\text{gh}}^2}{n} - \frac{2Y^2}{nk(k-1)} )</td>
</tr>
<tr>
<td>Among lines</td>
<td>( k - 1 )</td>
<td>( \sum_{g=1}^{k} 4 \left( \frac{k}{2} \right) Y_{\text{gh}}^2 - \frac{2Y^2}{nk^2 (k-2)} )</td>
</tr>
<tr>
<td>Within lines</td>
<td>( \frac{k(k-3)}{2} )</td>
<td>Subtraction</td>
</tr>
<tr>
<td>Error</td>
<td>( (n - 1)(m - 1) )</td>
<td>( \sum_{i=1}^{n} \sum_{g&lt;h=2}^{k} Y_{i\text{gh}}^2 - \sum_{i=1}^{n} \frac{2Y_i^2}{k(k-1)} )</td>
</tr>
<tr>
<td></td>
<td>( \frac{nk(k-1)}{2} - 1 )</td>
<td>( \sum_{i=1}^{n} \sum_{g&lt;h=2}^{k} Y_{i\text{gh}}^2 - \frac{2Y^2}{nk(k-1)} )</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>( \sum_{i=1}^{n} \sum_{g&lt;h=2}^{k} Y_{i\text{gh}}^2 - \frac{2Y^2}{nk(k-1)} )</td>
</tr>
</tbody>
</table>
The expectation of the replicate, variety and error mean squares are \( \sigma_r^2 + \frac{k(k-1)}{2} \sigma_t^2, \sigma_v^2 + n\sigma_t^2, \) and \( \sigma_e^2 \), respectively, which follow from the randomized complete block analysis. Using the regression equation (5.2) for \( y_{gh} \), the various totals are:

\[
Y_{gh} = n\mu + \sum_{i=1}^{n} r_i + n \left\{ l_g + l_h + (lt)_{gh} \right\} + \sum_{i=1}^{n} e_{igh} \tag{5.3}
\]

\[
Y_{g} = n(k-1)\mu + (k-1) \sum_{i=1}^{n} r_i + n(k-1)l_g
\]

\[
+ n \sum_{h=1}^{k} \left\{ l_h + (lt)_{gh} \right\} + \sum_{i=1}^{n} \sum_{h=1}^{k} e_{igh} \tag{5.4}
\]

\[
Y_{..} = \frac{k(k-1)}{2} \left\{ n\mu + \sum_{i=1}^{n} r_i \right\} + n(k-1) \sum_{g=1}^{k} l_g + \sum_{g<h=2}^{k} \left( \sum_{i=1}^{n} (lt)_{gh} \right) + \sum_{i=1}^{n} \sum_{g<h=2}^{k} e_{igh} \tag{5.5}
\]

The expectation of the sum of squares among the \( k(k-1)/2 \) crosses is

\[
\left\{ \frac{k(k-1)}{2} - 1 \right\} \left\{ \sigma_r^2 + n\sigma_t^2 \right\} + n(k-1) (k-2) \sigma_t^2. \tag{5.6}
\]

The preceding expectations may also be obtained in the following sum of squares

\[
E \left[ \sum_{g<h=2}^{k} \frac{k(k-1)}{2} \frac{Y_{gh} - Y_{..}}{nk^2(k-1)^2} \right]^2 \tag{5.7}
\]

with \( \frac{k(k-1)}{2} - 1 = \frac{(k+1) (k-2)}{2} \) degrees of freedom.

The expectation of the line sum of squares is

\[
E \left[ \sum_{j=1}^{k} \frac{k}{4} \left( \frac{k-1}{2} Y_{..} - Y_{..} \right) \right] = \left( k-1 \right) \left\{ \sigma_r^2 + n\sigma_t^2 + n(k-2) \sigma_t^2 \right\} \tag{5.8}
\]

with \( k-1 \) degrees of freedom. The divisor of equation (5.8) is \( nk^2 (k-2)/4 \) since the \( n(k-2) \) quantities \( l_g \) are
multiplied by \( k/2 \). The within line sum of squares is obtained by subtraction and has the expectation

\[
\frac{k(k-3)}{2} \left\{ \sigma^2_z + n\sigma_{1_z}^2 \right\}
\]

(5.9)

with \( k(k-3)/2 \) degrees of freedom.

For a paper written by Sprague and Tatum (17) W. G. Cochran had derived formulas for specific and general combining ability from an experiment such as that described above. The derivation of these estimates as indicated by Cochran follows.

6. Variance for general combining ability.

The yield for any particular plot is represented by the equation for \( y_{i_gh} \) as given previously. Now if the \( k \) lines are considered as the population it may be assumed that

\[
\sum_{h=1}^{k} (lt)_{gh} = 0 \quad \text{and} \quad \sum_{h=1}^{k} l_h = -l_g .
\]

Making use of these assumptions, the line and grand totals become

\[
Y_{..g} = (k - 1) (n\mu + \sum_{i=1}^{n} r_i) + n(k - 2)l_g
\]

\[
+ \sum_{i=1}^{n} \sum_{h=1}^{k} e_{1gh}
\]

(6.1)

\[
Y_{...} = \frac{k(k-1)}{2} (n\mu + \sum_{i=1}^{n} r_i) + \sum_{i=1}^{n} \sum_{g<h=2}^{k} e_{1gh}
\]

\[
= \frac{k}{2} \sum_{g=1}^{k} Y_{..g}
\]

(6.2)

The difference, \( \frac{k}{2} Y_{..g} - Y_{...} \), squared for any particular line \( g \) has the expectation

\[
\frac{n^2(k-2)^2}{4} k^2\sigma_{1_z}^2 + \frac{nk(k-1)(k-2)}{4} \sigma_z^2.
\]

(6.3)
Therefore,
\[ \hat{\sigma}_g^2 = \frac{(k - 1)}{k(k - 2)} \left\{ \frac{4}{n^2k(k - 1)(k - 2)} \sum_{h \neq g} \left( \frac{k - 2}{n(k - 2)(k - 3)} \right) \right\} \]
which is a form of the formula obtained by Cochran.

7. Variance for specific combining ability.

Using the equation for \( y_{lgh} \), the sum of squares for the effect of a particular line \( g \) crossed on the other \( (k - 1) \) lines has the expectation
\[ E \left[ \sum_{h=1}^{k} \left( \frac{(k - 2)Y_{gh} - Y_g - Y_h + \frac{2Y_{gh}}{k - 1}}{n(k - 2)(k - 3)} \right)^2 \right] = (k - 2) \left\{ \frac{\sigma^2}{n} + n\sigma^2 \right\} \]
with \( (k - 2) \) degrees of freedom. If the \( m_l \) is assumed to be zero, then the expectation of the above sum of squares was incorrectly given by Sprague and Tatum (17) as
\[ n(k - 2) \left\{ \frac{\sigma^2}{n} + \frac{(k - 1)}{(k - 3)} \sigma^2 \right\} \]

The correct expectation has been ascertained to be
\[ n(k - 2) \left\{ \frac{\sigma^2}{n} + \frac{(k - 1)}{(k - 3)} \sigma^2 \right\} \]
Sprague and Tatum (17) illustrate the use of these formulas with an example on corn hybrids involving \( k = 10 \) inbred lines or 45 single crosses of corn.

Considering the \( (k - 1) \) lines crossed with the line \( g \) as a sample of lines, the variance for the mean of the \( g \)th line is
\[ E \left[ \frac{Y_g}{n(k - 1)} \right]^2 - \left[ E \frac{Y_g}{n(k - 1)} \right]^2 = \frac{\sigma^2}{n(k - 1)} + \frac{\sigma^2}{(k - 1)} \]

since
\[ \sum_{i=1}^{n} r_i = 0 \quad \text{and} \quad \sum_{h=1}^{k} I_h = -1 \quad \text{for} \quad g \]

\( ^{5} \)By written correspondence with W. G. Cochran.
8. **Expectation of the mean squares for unequal numbers in each cell.**

The weighted squares of means analysis for an experiment of \( i = 1, \ldots, n_j \leq n \) replicates with \( j = 1, \ldots, m_i \leq m \) varieties in the \( i \)th replicate and \( p_{ij} \) sampling units in the \( j \)th variety in the \( i \)th replicate or in the \( ij \)th subplot is given below:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicates</td>
<td>( n - 1 )</td>
<td>( \sum_{i=1}^{n} \frac{Y_{i.}^2}{p_{i.}} - \frac{Y_{..}^2}{p_{..}} )</td>
<td>MS(R)</td>
</tr>
<tr>
<td>Varieties or treatments</td>
<td>( m - 1 )</td>
<td>( \sum_{j=1}^{m} \frac{Y_{j..}^2}{p_{j.}} - \frac{Y_{..}^2}{p_{..}} )</td>
<td>MS(V)</td>
</tr>
<tr>
<td>Experimental error</td>
<td>( \sum_{i=1}^{n} \frac{m_i}{m} \frac{Y_{i..}^2}{p_{i.}} - \frac{Y_{..}^2}{p_{..}} )</td>
<td></td>
<td>MS(RV)</td>
</tr>
<tr>
<td>Sampling error</td>
<td>( \sum_{i=1}^{n} m_i )</td>
<td>( \sum_{i=1}^{n} \left{ \frac{p_{ij}}{p_{..}} \sum_{k=1}^{m} y_{ijk}^2 \right} - \frac{Y_{..}^2}{p_{..}} )</td>
<td>MS(E)</td>
</tr>
<tr>
<td>Total</td>
<td>( p_{..} - 1 )</td>
<td>( \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{m_i}{m} \frac{p_{ij}}{p_{..}} \sum_{k=1}^{m} y_{ijk}^2 - \frac{Y_{..}^2}{p_{..}} )</td>
<td></td>
</tr>
</tbody>
</table>

In the above analysis

\[
Y_{i..} = \sum_{j=1}^{m_i} \sum_{k=1}^{p_{ij}} y_{ijk}, \tag{8.1}
\]
\[ Y_{ijk} = \frac{n_j}{\sum_{i=1}^{n_j} \sum_{k=1}^{p_{ij}} y_{ijk}}, \]  

(8.2)

grand total = \[ Y\ldots = \sum_{i=1}^{n} \sum_{j=1}^{m_{ij}} \sum_{k=1}^{p_{ij}} y_{ijk}, \]  

(8.3)

\[ p\ldots = \text{total number of all units in the experiment} \]

\[ = \sum_{i=1}^{n} m_{ij} p_{ij}, \]  

(8.4)

\[ p_{i.} = \sum_{j=1}^{m_{ij}} p_{ij}, \]  

(8.5)

\[ p_{.j} = \sum_{i=1}^{n_{ij}} p_{ij}, \]  

(8.6)

and \( y_{ijk} \) is the yield of a sampling unit. If it is assumed that the effects contributing to the yield of an individual unit are independent and additive, the linear equation

\[ y_{ijk} = \mu + r_i + v_j + (rv)_{ij} + e_{ijk}, \]  

(8.7)

may be used to express the yield of an individual sampling unit where \( \mu \) is the population mean, \( r_i \) is an effect common to the \( i \)th replicate, \( v_j \) an effect common to the \( j \)th variety, \( (rv)_{ij} \) an effect common to the \( j \)th variety in the \( i \)th replicate, \( e_{ijk} \) an effect common to every unit, \( i = 1, 2, \ldots, n_j \leq n \), \( j = 1, 2, \ldots, m_{ij} \leq m \), and \( k = 1, 2, \ldots, p_{ij} \).

The expected value for the replicate mean square is

\[ \sigma^2 + \frac{1}{n-1} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m_{ij}} p_{ij}^2 \left[ \frac{1}{p_{i.}} - \frac{1}{p_{.j}} \right] \sigma^2_{rv} \right\} \]

\[ + \left[ p_{.} - \frac{1}{p_{.}} \sum_{i=1}^{n} p_{i.} \right] \sigma^2_r + \left[ \sum_{i=1}^{n} \sum_{j=1}^{m_{ij}} p_{ij}^2 \right] \frac{m_{ij}}{p_{i.}} \]

\[ - \frac{1}{p_{.}} \sum_{j=1}^{p_{.j}} p_{.j} \right\} \sigma^2_v \]  

(8.8)

For the case where \( n_j = n \), \( m_{ij} = m \), \( p_{ij} = p \), and \( p\ldots = mnp \), the replicate mean square has the expectation

\[ \sigma^2 + p \sigma^2_{rv} + m p \sigma^2_v. \]  

(8.9)
The expected value for the variety mean square is

\[ \sigma_v^2 \left(1 + \frac{1}{n} \sum_{j=1}^{1} \sum_{i=1}^{1} p_{ij}^2 \right) \]

\[ + \left[ \frac{1}{p_{..}} - \frac{1}{p_{..}} \sum_{j=1}^{1} p_{j} \right] \sigma_r^2 \]

\[ + \left[ \frac{1}{p_{..}} - \frac{1}{p_{..}} \sum_{j=1}^{1} p_{j} \right] \sigma_r^2 \]

\[ = \sigma_r^2 \left(1 + \frac{1}{n} \sum_{j=1}^{1} \sum_{i=1}^{1} p_{ij}^2 \right) \]

\[ + \left[ \frac{1}{p_{..}} - \frac{1}{p_{..}} \sum_{j=1}^{1} p_{j} \right] \sigma_r^2 \]

For the case of equal numbers for all subclasses the expected value for the variety mean square is

\[ \sigma_v^2 + p \sigma_r^2 + n p \sigma_v^2. \]  

The expected value for the replicate by variety mean square is

\[ \sigma_r^2 \left(1 + \frac{1}{n} \sum_{j=1}^{1} \sum_{i=1}^{1} p_{ij}^2 \right) \]

\[ + \left[ \frac{1}{p_{..}} - \frac{1}{p_{..}} \sum_{j=1}^{1} p_{j} \right] \sigma_r^2 \]

\[ + \left[ \frac{1}{p_{..}} - \frac{1}{p_{..}} \sum_{j=1}^{1} p_{j} \right] \sigma_r^2 \]

For the case of equal numbers in all categories, the replicate by variety mean square has the expectation

\[ \sigma_r^2 \left(1 + \frac{1}{n} \sum_{j=1}^{1} \sum_{i=1}^{1} p_{ij}^2 \right) \]

\[ + \left[ \frac{1}{p_{..}} - \frac{1}{p_{..}} \sum_{j=1}^{1} p_{j} \right] \sigma_r^2 \]

For the more simple case where

\[ y_{ij} = \mu + r_i + v_j + (rv)_{ij}, \]  

9. **Expectation of the mean squares for one or zero individuals in each cell.**

For the more simple case where

\[ y_{ij} = \mu + r_i + v_j + (rv)_{ij}, \]  

\[ (9.1) \]
the expected value for the replicate mean square is
\[
\frac{1}{n-1} \sum_{i=1}^{n} \frac{Y_{i}^2}{m_i} \left( \frac{1}{m} \sum_{j=1}^{m} Y_{i,j}^2 \right) - \frac{1}{m} \sum_{i=1}^{n} m_i^2 \sigma_r^2 + \frac{1}{n-1} \left\{ \sum_{i=1}^{m} \frac{m}{m} \sum_{j=1}^{n} \frac{n_j}{n} \right\} \sigma_r^2 \] (9.2)

which is equal to
\[
\sigma_{rv}^2 + m\sigma_r^2 \] (9.3)

for all \( m_i \) equal, all \( n_j \) equal and \( m = n = mn \).
Likewise the expected mean square for varieties is
\[
\frac{1}{m-1} \sum_{j=1}^{m} \frac{Y_{j}^2}{n_j} \left( \frac{1}{n} \sum_{i=1}^{n} y_{i,j}^2 \right) - \frac{1}{n} \sum_{j=1}^{m} n_j^2 \sigma_r^2 + \frac{1}{m-1} \left\{ \sum_{j=1}^{m} \frac{n}{n} \sum_{i=1}^{m} \frac{m_i}{m} \right\} \sigma_r^2 \] (9.4)

which for the orthogonal case \( (m = n = mn, n_j = n, m_i = m) \) is equal to
\[
\sigma_{rv}^2 + n\sigma_r^2 \] (9.5)

The error mean square has the expectation
\[
\frac{1}{m - m - n + 1} \sum_{i=1}^{m} \frac{m_i}{i} \left( \frac{1}{n} \sum_{j=1}^{n} y_{i,j}^2 \right) \left( \frac{1}{m} \sum_{j=1}^{m} \frac{Y_{j}^2}{n_j} \right) - \frac{1}{n} \sum_{i=1}^{m} m_i^2 \sigma_r^2 + \frac{1}{m - m - n + 1} \left\{ \frac{1}{n} \sum_{i=1}^{m} m_i^2 - m \right\} \sigma_r^2 \] (9.6)

which for the orthogonal case is equal to \( \sigma_{rv}^2 \).

Since \( MS(R), MS(V) \) and \( MS(RV) \) each contain estimates of the variance \( \sigma_{rv}^2, \sigma_r^2 \) and \( \sigma_r^2 \), it is necessary to use all three mean squares to obtain estimates for the various variance components.

Although the above method provides unbiased estimates for the components of variance, nothing is known about the efficiency of these estimates as compared to an efficient
method, for example, the method of maximum likelihood. If the numbers in the subclasses are greatly different and if the degrees of freedom for the various mean squares are small, then the efficiency of variance components from the weighted squares of means analysis is probably low.

In replicated yield tests on hybrid corn the number of varieties is usually large and the number of replicates per variety does not vary greatly. Therefore, the unbiased estimates of the variance components from the weighted squares of means analysis are probably suitable for most purposes. This is especially true for the replicate by variety variance component and for the variety variance component. The replicate variance component, which is of least interest, is estimated the least efficiently, but even so it may be fairly efficient since the number of varieties per replicate will usually be nearly equal.

OTHER METHODS OF ANALYSIS FOR THE TWO-WAY CLASSIFICATION WITH ONE OR ZERO INDIVIDUALS IN THE SUBCLASSES

10. The method of fitting constants.

For the hybrid corn experiments, it is necessary to consider only the case of missing plot yields from randomized complete block experiments. The \( i \)th replicate will give yields on \( m_i \) varieties, and the \( j \)th variety total yield is obtained from the \( n_j \) replicates in which the variety was present. The yield for the \( j \)th variety in the \( i \)th replicate may be expressed by the linear equation

\[
Y_{ij} = \mu + r_i + v_j + (rv)_{ij},
\]

where \( \mu \) is the mean yield over all plots, \( r_i \) is the effect common to the \( i \)th replicate, \( v_j \) is the effect common to the \( j \)th variety, and \( (rv)_{ij} \) is an effect common to \( ij \)th plot; \( i = 1, 2, \ldots, n \) and \( j = 1, 2, \ldots, m \). For the randomized complete block experiment with no missing plots \( n_j = n \) and \( m_i = m \).

The normal equation for the mean is

\[
\sum_{i=1}^{n} m_i \mu + m_i r_i + \ldots + m_n r_n + n_i v_i + \ldots + n_m v_m = Y \ldots .
\]

For the variety effects the normal equations are (for the \( j \)th variety)

\[
n_j \mu + n_j v_j + r_i + \ldots + r_n = Y_{.j} = \sum_{i=1}^{n_j} y_{ij} .
\]
The normal equations for the replicate effects are (for
the \( i \)th replicate)

\[
m_i \mu + m_i r + v_1 + \ldots + v_m = Y_{i1} = \sum_{j=1}^{m} y_{ij}.
\]

(10.4)

For this particular case, there are \( mn \) total subclasses
or plots with \( \sum_{i=1}^{n} (m - m_i) = \sum_{j=1}^{m} (n - n_j) \) missing yields.

For purposes of simplicity and with no loss in generality let
variety 1 be the variety with yields from the largest num­
ber of replicates \( n_1 \); let variety 2 be the one with yields from
the next largest number of replicates \( n_2 \), etc. Thus \( n \geq n_1
\geq n_2 \geq \ldots \geq n_m > 0 \). Likewise let replicate 1 be the one
with the largest number of varieties \( m_1 \), replicate 2 with the
next largest number of varieties \( m_2 \), etc. Thus \( m \geq m_1 \geq m_2
\geq \ldots \geq m_m > 0 \). The two-way classification for variety
yields could be represented as follows:

<table>
<thead>
<tr>
<th>Varieties</th>
<th>Replicates</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 ... i</td>
<td>Y_{11}</td>
</tr>
<tr>
<td>1</td>
<td>y_{11}</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>y_{12}</td>
<td>\ldots</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>y_{1j}</td>
<td>\ldots</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>y_{1m}</td>
<td></td>
</tr>
<tr>
<td>Total yield</td>
<td>Y_{1}</td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>m _1</td>
<td></td>
</tr>
</tbody>
</table>

In order to obtain a unique solution for the normal
equations, it is necessary to impose restrictions. Two re­
strictions which yield a unique solution are

\[
m_i \geq m_1 \geq m_2 \geq \ldots \geq m_m > 0, \quad n \geq n_1 \geq n_2 \geq \ldots \geq n_m > 0.
\]
These relations do not appear as such in the normal equations. Now, \( \sum_{i=1}^{n} m_i r_i \) may be written as

\[
\sum_{i=1}^{n} \left[ m - (m - m_i) \right] r_i = - \sum_{i=1}^{n} (m - m_i) r_i \tag{10.5}
\]

since \( \sum_{i=1}^{n} r_i = 0. \)

Likewise,

\[
\sum_{j=1}^{m} n_j v_j = - \sum_{j=1}^{m} (n - n_j) v_j,
\]

\[
\sum_{i=1}^{n} m_i r_i = - \sum_{i=1}^{n} (m - m_i) r_i - \sum_{i=n_j+1}^{m} m_i r_i, \tag{10.7}
\]

and

\[
\sum_{j=1}^{m} n_j v_j = - \sum_{j=1}^{m} (n - n_j) v_j - \sum_{j=m_i+1}^{n_j} n_j v_j. \tag{10.8}
\]

Estimates of the constants in the normal equations are required. \( \mu \) is obtained from the normal equation for the mean, thus

\[
\mu = \frac{1}{m} \left\{ \sum_{i=1}^{n} r_i m_i - \sum_{j=1}^{m} n_j v_j \right\}
\]

\[
= \frac{1}{m} \left\{ \sum_{i=1}^{n} (m - m_i) r_i + \sum_{j=1}^{m} (n - n_j) v_j \right\}. \tag{10.9}
\]

Also,

\[
v_j = \frac{Y_{.j}}{n_j} - \frac{1}{m} \left\{ \sum_{i=1}^{n} (m - m_i) r_i + \sum_{j=1}^{m} (n - n_j) v_j \right\}
\]

\[
+ \frac{1}{n_j} \sum_{i=n_j+1}^{n} r_i \tag{10.10}
\]

and
Using the regression notation of Yates (22) the reduction in the total sum of squares due to fitting the constants $\mu$, $r_i$, and $v_j$ is the error or replicate by variety sum of squares. The expectation of this quantity is

$$E \left[ \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}^2 - \left\{ \mu Y.. + \sum_{i=1}^{n} r_i Y_{i..} + \sum_{j=1}^{m} v_j Y_{.j} \right\} \right] = (m - m - n + 1) \sigma_{r,v}^2$$

with $m - m - n + 1$ degrees of freedom.

The difference in the reduction of the total sum of squares due to fitting the constants $\mu$, $r_i$, and $v_j$ and the reduction due to fitting the constants $\mu$ and $r_i$ only is the sum of squares for varieties by the method of fitting constants. Symbolically this sum of squares is

$$\mu Y.. + \sum_{i=1}^{n} r_i Y_{i..} + \sum_{j=1}^{m} v_j Y_{.j} - \sum_{i=1}^{n} \frac{Y_{i..}^2}{m_i}$$

and has the expectation

$$(m - 1) \sigma_{r,v}^2 + (m - n) \sigma_r^2 = E \left[ \sum_{j=1}^{n} v_j Y_{.j} \right]$$

with $(m - 1)$ degrees of freedom.

Likewise, the replicate sum of squares by the method of fitting constants has the expectation

$$E \left[ \mu Y.. + \sum_{i=1}^{n} r_i Y_{i..} + \sum_{j=1}^{m} v_j Y_{.j} - \sum_{j=1}^{m} \frac{Y_{.j}^2}{n_j} \right]$$

$$= (n - 1) \sigma_{r,v}^2 + (m - m) \sigma_r^2 = E \left[ \sum_{i=1}^{n} r_i Y_{i..} \right]$$

with $(n - 1)$ degrees of freedom.

The total sum of squares has the expectation

$$E \left[ \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}^2 - \frac{Y..^2}{m.} \right] + (m - 1) \sigma_{r,v}^2$$

$$+ (m - \frac{1}{m} \sum_{j=1}^{n} n_j) \sigma_r^2 + (m - \frac{1}{m} \sum_{i=1}^{n} m_i) \sigma_v^2$$

(10.16)
with \((m - 1)\) degrees of freedom. As a check the error and variety sum of squares by the method of fitting constants plus the replicate sum of squares by the weighted squares of means analysis should equal the total.

11. **Missing plot formula and expectation of mean squares for \(k\) values missing in one group of the other classification.**

Yates (21) has described a method by which the missing or incomplete results may be calculated and the resulting tests of significance made. The missing values are computed in such a manner as to make the error sum of squares a minimum. The expectation of the variety and replicate mean squares from a randomized complete block experiment with the calculated missing plot values included in the analysis (21) will depend upon the location of the missing values in the experiment. The coefficients for the variety and replicate variance components for the respective mean squares will be \(n\) = number of replicates and \(m\) = number of varieties, respectively, and are the same coefficients as for the orthogonal case where no plot yields are missing or incomplete. However, the coefficient of error variance component in the variety or replicate mean square will *not* be the same as for the orthogonal case. The coefficient for the error variance component in the error mean square will be unity if the degrees of freedom are taken equal to \((n - 1)(m - 1) - k\), where \(k\) is the number of calculated yields in the experiment.

The case of \(k < m\) varietal yields missing in one of the replicates is considered first. The results are applicable also to the case of \(k < n\) missing yields for one variety simply by a change of symbols \(n\) and \(m\) and \(\sigma_r^2\) and \(\sigma^2\). The table of yields, \(y_{i\ell}\), with the calculated missing values, \(x_{i\ell}\) \((\ell = 1, 2, \ldots, k)\), may be represented as follows for a two-way classification or for varieties and replicates in particular:
The analysis of the above experiment is performed in the same manner as for any $m$ by $n$ classification with no missing yields except that the error degrees of freedom will be $k$ less than for the orthogonal case. The error sum of squares is

$$
E = \sum_{t=1}^{k} x_{it}^2 + \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij}^2 - \frac{1}{n} \left\{ \sum_{t=1}^{k} (Y_{..} + x_{it})^2 \right\} - \frac{1}{m} \left\{ \left( Y_{..} + \sum_{t=1}^{k} x_{it})^2 + Y_{2..}^2 + \ldots + Y_{n..}^2 \right\} \\
+ \frac{1}{mn} \left\{ Y_{..} + \sum_{t=1}^{k} x_{it} \right\}^2.
$$

Upon minimizing this function, $k$ linear equations in $x_{11}, x_{12}, \ldots, x_{1k}$ are obtained. The linear equations are

$$
x_{it} (m - 1)(n - 1) - (n - 1) \left( \sum_{t=2}^{k} x_{it} = nY_{1..} + mY_{..} - Y_{..} \right) (11.2)
$$

$$
x_{i2} (m - 1)(n - 1) - (n - 1) \left( \sum_{t=1}^{k} x_{it} = nY_{1..} + mY_{..} - Y_{..} \right) (11.3)
$$
\[ -(n-1)^{k-1} \sum_{t=1}^{x_{it} + x_{ik} (m-1)(n-1)} = nY_1 + mY_{m-1}Y \ldots \] (11.4)

Although the above equations are easily solved by iterative methods (21) the solutions for the calculated yields for the missing plots are also easy to obtain for the case of \( k \) missing values in one replicate, thus

\[
x_{i1} = \frac{1}{(m-k)(n-1)} \left\{ nY_1 - Y \ldots + (m-k)Y_1 \right\} + \sum_{t=1}^{k} Y_{i,t}
\] (11.5)

\[
x_{i2} = \frac{1}{(m-k)(n-1)} \left\{ nY_1 - Y \ldots + (m-k)Y_2 \right\} + \sum_{t=1}^{k} Y_{i,t}
\] (11.6)

\[
\ldots
\]

\[
x_{ik} = \frac{1}{(m-k)(n-1)} \left\{ nY_1 - Y \ldots + (m-k)Y_k \right\} + \sum_{t=1}^{k} Y_{i,t}
\] (11.7)

Now the calculated value \( x_{it} \) \((t = 1, 2, \ldots, k)\) is equal to \( \mu + r_i + v_j + f (rv_{ij}) \), where \( f (rv_{ij}) \) is some function of the errors in the experiment. The calculated plot yields will be correlated to some extent with the \( rv_{ij} \) error effect in the remainder of the experiment and with each other. The contributions of the individual errors \( rv_{ij} \) to the calculated yields for the \( x_{it} \) and for the \( x_{it} \) with \( x_{1w} \) \((t = 1, \ldots, k)\) are given in table 4.

The sum \( n/ (n-1)(m-k) \) (see table 4) multiplied by 2 gives the covariance term of the error variance component for any two calculated yields. The covariance of the calculated missing plot yield with the errors in any variety or
<table>
<thead>
<tr>
<th>Replicates</th>
<th>Varieties</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(x_{it})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1/(n-1)</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>1/(n-1)</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(x_{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(x_{it} \text{ with } x_{iw})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
replicate is \(2(1) = 2\). The sum of squares of the coefficients for any \(x_{1t}\) is

\[
\frac{m + n - k}{(n-1)(m-k)}. \tag{11.8}
\]

With these results and expressing the yield of an individual plot as \(y_{ij} = \mu + r_i + v_j + rv_{ij}\), it is possible to express the expectation for the mean squares in terms of the variance components \(\sigma^2_{rv}\), \(\sigma^2_r\) and \(\sigma^2_v\).

The correction term has the expectation

\[
mn\mu^2 + m\sigma^2_r + n\sigma^2_v + \sigma^2_{rv} \left\{ 1 + \frac{k}{(n-1)(m-k)} \right\}. \tag{11.9}
\]

The replicate sum of squares has the expectation

\[
m(n-1)\sigma^2_r + \sigma^2_{rv} (n-1) \left\{ 1 + \frac{k}{(m-k)(n-1)} \right\} \tag{11.10}
\]

with \((n-1)\) degrees of freedom.

The variety mean square has the expectation

\[
n\sigma^2_v + \sigma^2_{rv} \left\{ 1 + \frac{k}{(n-1)(m-1)} \right\}. \tag{11.11}
\]

The error sum of squares has the expectation

\[
\left\{ mn - k - m - n + 1 \right\} \sigma^2_{rv} \tag{11.12}
\]

with \((mn - k) - m - n + 1\) degrees of freedom.

12. **Missing plot formula and expectation of mean squares for \(k\) missing values for different entries in each of the two-way classifications.**

In the previous section the expectations of the mean squares from a two-way classification, or a randomized complete block design in particular, were obtained for the case of \(k\) calculated yields for missing or incomplete yields which were in one group of one of the classifications. The case in which the yields for different varieties are missing in different replicates will now be considered. The table of yields with the calculated missing yields, \(x_{it} (t = 1, 2, \ldots, k)\), for a two-way classification, or for a randomized complete block in particular, may be represented as follows:
The arrangement of the yields in this manner causes no loss in generality. The analysis of variance of the yields including the calculated yields is performed in the same manner as for the orthogonal randomized complete block except that $k$ is subtracted from the error degrees of freedom (21).

Upon minimizing the error sum of squares and solving the $k$ resulting equations, the calculated values for the missing or incomplete plot yields are

$$
x_{tt} = \frac{1}{(mn-m-n)(mn-m-n+k)} \left\{ (mn-m-n+k) \left( nY_{tt} + mY_{..} \right) - m \sum_{t=1}^{k} Y_{..} - n \sum_{t=1}^{k} Y_{..} \left( mn-m-n \right) Y_{..} \right\}
$$

(12.1)

The replicate, variety and error sums of squares have the expectations

$$(n-1) \left[ m\sigma_r^2 + \sigma_{rv}^2 \left\{ 1 + \frac{k(n-1)(mn-m-n) + nk(k-1)}{(n-1)(mn-m-n+k)(mn-m-n)} \right\} \right] , \quad (12.2)$$

<table>
<thead>
<tr>
<th>Replicates</th>
<th>Varieties</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>$x_{11}$</td>
<td>$y_{12}$</td>
</tr>
<tr>
<td>2</td>
<td>$y_{21}$</td>
<td>$x_{22}$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$k$</td>
<td>$y_{k1}$</td>
<td>$y_{k2}$</td>
</tr>
<tr>
<td>$k+1$</td>
<td>$y_{k+1,1}$</td>
<td>$y_{k+1,2}$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$n$</td>
<td>$y_{n1}$</td>
<td>$y_{n2}$</td>
</tr>
<tr>
<td>Totals</td>
<td>$Y_{..} + x_{11}$</td>
<td>$Y_{..} + x_{22}$</td>
</tr>
</tbody>
</table>
\[ (m - 1) \left[ n\sigma^2 + \sigma_{r^2} \left\{ 1 + \frac{k(m - 1)(mn - m - n) + mk(k - 1)}{(m - 1)(mn - m - n + k)(mn - m - n)} \right\} \right], \quad (12.3) \]

and

\[ \sigma_{r^2} \left\{ mn - k - m - n + 1 \right\} \quad (12.4) \]

with \((n - 1), (m - 1)\) and \((mn - k - m - n + 1)\) degrees of freedom, respectively.

In this and the preceding section missing plot formulas were developed for two special cases and the expectations for the resulting mean squares obtained. Federer (6) has given missing plot formulas for \(p\) missing plots in each of the \(n\) replicates with no variety missing more than once, for \(p_1\) varieties missing in one replicate and \(p_2\) different varieties missing in another replicate, and for \(p_1, p_2\) and \(p_3\) different varieties missing in three replicates. No generalized missing plot formulas were presented but the general solution was indicated. From a practical standpoint it appears that any generalized formula would be too complicated for general usage and the iterative method of Yates (21) would be used instead.


Although the method of fitting constants for a two-way classification with disproportionate subclass numbers is usually tedious, it may be applied to a randomized block experiment with missing plots by a simple dodge. Following the results obtained by Yates (21), the procedure for obtaining the method of fitting constants analysis for a randomized block experiment with missing plot values is:

1. Missing plot values are computed so as to make the error sum of squares a minimum [Yates (21), or sections 11 and 12].

2. With the missing plot values included in the table of yields the analysis of variance is computed in the same manner as for a randomized block experiment with no missing yields except that the error degrees of freedom are reduced by the number of calculated yields. The estimate of the error sum of squares is a minimum and will be the same as for the method of fitting constants.
3. The replicate sum of squares is computed by the method of weighted squares of means (sections 8 and 9).

\[
\frac{\sum_{i=1}^{n} m_i}{Y_{i.}^2 - \frac{Y_{..}^2}{m}}. \tag{13.1}
\]

4. The total sum of squares of the observed values is obtained from the formula

\[
\frac{\sum_{i=1}^{n} m_i}{Y_{ij}^2 - \frac{Y_{..}^2}{m}}. \tag{13.2}
\]

5. The variety sum of squares for the method of fitting constants is obtained by subtraction of the replicate (step 3) and the error (step 2) sums of squares from the total (step 4).

The analysis of variance (for a randomized complete block experiment with \(k\) missing plots) by the method of fitting constants is:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Average Value of Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicates ((n-1))</td>
<td>(\sigma_{rv}^2 + \frac{\sigma_r^2}{n-1} \left{ \frac{m}{m} - \frac{1}{m} \right} \sum_{i=1}^{m} m_i^2 \right} \bigg) + \frac{\sigma_v^2}{n-1} \left{ \frac{n}{m} - \frac{1}{m} \right} \sum_{j=1}^{m} n_j^2 \right} \bigg)</td>
<td></td>
</tr>
<tr>
<td>Varieties ((m-1))</td>
<td>(\sigma_{rv}^2 + \left( \frac{m-n}{m-1} \right) \sigma_r^2 )</td>
<td></td>
</tr>
<tr>
<td>Error ((m-1)(n-1)-k)</td>
<td>(\sigma_{rv}^2 )</td>
<td></td>
</tr>
</tbody>
</table>
The replicate sum of squares by the method of fitting constants may be obtained in a manner similar to that just described for varieties. All these results follow from the theory given by Yates (21) and from the expectations of the mean squares given in the preceding sections.

Another procedure (21) for obtaining the method of fitting constants analysis would be the following:

1. Compute the missing plot values.
2. With the missing values inserted in the table of yields the results are orthogonal and the estimation of the regression coefficients $\mu$, $v$, and $r$ proceeds in the same manner as the orthogonal case.
3. With these estimates of $\mu$, $v$, and $r$ the method of analysis proceeds in the same manner as that ordinarily used for obtaining sums of squares by the method of fitting constants (see section 10).

The first method given above is quite simple and straightforward and is preferred. It was the one used for some of the experiments summarized. This method of computing the analysis of variance is designated as method of analysis number 3.

EXPERIMENTAL RESULTS AND DISCUSSION

GENERAL DESCRIPTION OF THE DATA

A number of yield trials are conducted annually in the corn improvement program of the Iowa Agricultural Experiment Station. One of the main objectives of these tests is to select the highest yielding corn strains for the various sections of Iowa. The entries in a yield trial may be single crosses, double crosses, top crosses, inbreds, open-pollinated varieties, synthetic varieties, or some other strain of corn, or they may consist of various combinations of the different types of corn strains. The field design for a yield test may be a randomized complete block, lattice, latin square, split plot or other. The yield trial may be planted at more than one place or location.

The data used in this study represent 302 corn yield trials with entries and designs varying as indicated above. These data were obtained from Dr. G. F. Sprague and Dr. J. L. Robinson of the Iowa Agricultural Experiment Station. The 232 experiments obtained from Dr. Sprague are designated as the Iowa Experimental Corn Yield Trials (IECYT). The 70 experiments obtained from Dr. Robinson are designated as the Iowa Corn Yield Tests (ICYT). The 302 experiments
were planted during the years 1940 to 1946. A description of the number of experiments representing the various categories is presented in table 5. Fifty-eight percent of the experiments were made up of double crosses. The experiments containing entries which were other than top crosses, single crosses or double crosses constitute only a small fraction (2.6 percent) of the total.

The relationship between yield and number of missing hills is approximately linear for five or less missing hills. Therefore, it is the practice at the Iowa Agricultural Experiment Station to adjust plot yields with five or less missing hills and to calculate the expected yields for the plots when there are more than five missing hills. The data from the ICYT group, and part of the experiments for 1943 and 1944 and all of the experiments for 1945 and 1946 of the IECYT were treated in this manner. In the remaining IECYT, the plot yields were adjusted for 1 to 19 missing hills. If all 20 hills were missing, missing plot formulas were used to compute the expected yields. Since the average experimental errors, for the years in which the missing plot values were calculated for plots having more than five missing hills and for the years in which the individual plot yields were adjusted by a linear relation for number of missing hills, are nearly identical [see Federer (6), tables 11 to 18], the two groups are considered as one.

### TABLE 5. SOURCE AND TYPE OF DATA AND NUMBER OF EXPERIMENTS OF CORN YIELD TRIALS.

<table>
<thead>
<tr>
<th>Year in which experiment was grown</th>
<th>Iowa experimental corn yield trials* Type of hybrid or strain chiefly concerned</th>
<th>Iowa corn yield tests ** Type of hybrid or strain chiefly concerned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top cross (No.)</td>
<td>Single cross (No.)</td>
</tr>
<tr>
<td>1940</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>1941</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>1942</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>1943</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>1944</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>1945</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>1946</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td>73</td>
</tr>
</tbody>
</table>

* The data for these experiments were obtained from G. F. Sprague.
** The data for these experiments were obtained from J. L. Robinson.
*** Other includes variety, synthetic, F2 and inbreds.
14. **Analysis of the data.**

The detailed analysis for each of the 302 experiments was presented by Federer (6) in tables 11 to 18 inclusive. The randomized block analysis was used on all experiments even though the majority of the designs were lattices. Due to the fact that some of the plot yields were lost due to floods, rodents, human errors, etc., all of the experiments could not be treated in the same manner. Three types of randomized block analyses were used depending upon the number and location of missing plots. For the case of no missing plots, the analysis was the same as that for the ordinary randomized complete block; this was designated as type of analysis number 1. Type of analysis number 2 was applied to those cases for which all missing plots were in one replicate or for which the missing plots were for a different variety or strain in a different replicate (see sections 11 and 12); for this method of analysis the missing plot values were calculated and substituted in the table of yields and then the experiment was analyzed in the same manner as number 1 except that the error degrees of freedom were reduced by the number of missing plots.

Type of analysis number 3 is the method of fitting constants as described in section 13, i.e., method number 2 is followed through to obtain the error sum of squares, which is a minimum; then the replicate sum of squares is obtained by the method of weighted squares of means (section 9); and the variety sum of squares for the method of fitting constants (section 10) is obtained by subtraction of the error and the replicate sums of squares from the total sum of squares of the observed yields. The replicate mean square for type of analysis number 3 is the one obtained from the weighted squares of means analysis (see sections 8 and 9).

The calculation of the variance components is straightforward for the three methods of analysis. The replicate variance components for type of analysis number 3 were obtained from the replicate mean square calculated by the method of fitting constants. The error mean square in all three analyses is an estimate of $\sigma_e^2$. The variety mean square for methods of analysis number 1, 2 and 3 has the expectation $\sigma_e^2 + n\sigma_e^2$, $\sigma_e^2 + n\sigma_e^2$ and $\sigma_e^2 + \frac{mn - k - n}{m - 1}\sigma_e^2$, respectively, where $n$ equals number of replicates, $m$ equals number of varieties, $k$ equals number of missing plots, and $c$ is the coefficient of $\sigma_e^2$ as described in sections 11 and 12.
The replicate mean squares are of a similar nature. The unbiased estimate of the ratio of the variety variance component, $\sigma_v^2$, to the error variance component, $\sigma_e^2$, is obtained from formulas (1.3), (1.4) and (1.5) for types of analysis number 1, number 2 and number 3, respectively.

15. Correlations among items in the analyses of variance.

Various correlations were obtained among the variables: number of varieties, number of replicates, error variance component, variety variance component, replicate variance component, and the mean of the experiment. The correlations among the above-listed items are given in table 6, along with the number of items used.

Considering the two groups of correlations for the IECYT and ICYT over the years 1940 through 1946, there was nothing of outstanding interest. The correlation between the error and replicate variance components is the largest. This might be expected since a more variable soil would give higher error and replicate mean squares. This certainly would be true if there were a soil gradient in one direction. The number of varieties in an experiment is related to some extent to the size of the experimental error. The number of replicates and varieties are each positively related to the size of the replicate variance component. The variety variance component is related slightly to the experimental mean, but the replicate and error variance components do not appear to have much if any relationship with the experimental mean.

Since none of the correlations from the two groups of data in table 6 are large, the majority of the variation in one of the variables is unaccounted for by the variation in the other. This is what was desired for the methods presented. High correlations among various items, for example, between number of varieties and error mean square and between error and variety variance components, would have made it necessary to modify the method of averaging ratios.

16. Frequency distribution of variance components and of the $\hat{a}_n$ from 302 corn experiments.

Frequency distributions of the error, variety and replicate variance components were obtained for the Iowa Experimental Corn Yield Trials and for the Iowa Corn Yield Tests. The patterns of the frequency distributions (see table 7) for both groups of experiments are very much alike. This might be expected since the two groups were planted in the same locality each year, and if heavy rains, say, af-
<table>
<thead>
<tr>
<th>Experimental group</th>
<th>Yr.</th>
<th>Variable</th>
<th>No. of varieties</th>
<th>No. of replicates</th>
<th>Error mean square</th>
<th>Variety variance component</th>
<th>Replicate variance component</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iowa experimental corn yield tests</td>
<td>1940</td>
<td>Error mean square</td>
<td>.218**</td>
<td>.088</td>
<td>.142*</td>
<td></td>
<td></td>
<td>232</td>
</tr>
<tr>
<td>to 1946</td>
<td></td>
<td>Variety variance component</td>
<td>.051</td>
<td>.165*</td>
<td></td>
<td>.021</td>
<td>.148*</td>
<td>232</td>
</tr>
<tr>
<td></td>
<td>1946</td>
<td>Replicate variance component</td>
<td>.116</td>
<td>.200**</td>
<td>.419**</td>
<td></td>
<td></td>
<td>232</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experiment mean</td>
<td></td>
<td></td>
<td>.073</td>
<td>.148*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iowa corn yield tests</td>
<td>1940</td>
<td>Error mean square</td>
<td>.045</td>
<td>.046</td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td>to 1946</td>
<td></td>
<td>Variety variance component</td>
<td>-.154</td>
<td>-.210</td>
<td>.083</td>
<td>-.006</td>
<td></td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>1946</td>
<td>Replicate variance component</td>
<td>.448**</td>
<td>.357**</td>
<td>.461**</td>
<td>-.006</td>
<td>.388**</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experiment mean</td>
<td></td>
<td></td>
<td>.168</td>
<td>.388**</td>
<td>-.006</td>
<td></td>
</tr>
</tbody>
</table>

* Significant at 5 percent level of probability.
** Significant at 1 percent level of probability.
<table>
<thead>
<tr>
<th>Variance component</th>
<th>Class interval</th>
<th>Replicate</th>
<th>ICYT</th>
<th>IECYT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>0-2.99</td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>3-5.99</td>
<td></td>
<td></td>
<td></td>
<td>302</td>
</tr>
<tr>
<td></td>
<td>6-8.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9-11.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12-14.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15-17.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18-20.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21-23.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24-25.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26-27.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28-29.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30-31.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 7. FREQUENCY DISTRIBUTION OF THE ERROR, VARIETY, AND REPLICATE VARIANCE COMPONENTS FROM 302 CORN YIELD TESTS.**
affected the stands in one group they were likely to affect the
stands of the other group.

The error and variety variance components give similar
frequency distributions both in numerical range and in shape.
The mean of the variety variance components from all ex-
periments, 8.294, is somewhat smaller than the mean, 12.794,
for the error variance component. The frequency distribu-
tion of the replicate variance components is much more
closely centered around its mean, 2.955, than are the other
two variance components; also, a larger proportion, 25/302,
of the replicate variance components are negative.

The frequency distribution of the $\hat{a}$'s is given in table 8.
Although the numerical value of the range for the unbiased
estimate of the ratio of the variety to the error variance
components is smaller (15.9) than either of the ranges (74
or 81) of the error or variety variance components, the range
relative to the mean (16.6) is larger than either of the lat-
ter two relative ranges (5.8 and 9.8), respectively. If the
variation is not related to the mean, then the variation in
the $\hat{a}$'s is smaller than for the error or variety variance com-
ponents.

The patterns of the distribution of the $\hat{a}$'s for top crosses,
single crosses and double crosses appear to be similar. How-
ever, the pattern for top crosses appears to be a little more
skewed than for the others, but this may be sampling vari-
ation resulting from the relatively small number of observa-
tions of the $\hat{a}$'s from top cross experiments. The frequency
distributions (table 8) for the IECYT and the ICYT double
crosses were alike in pattern and in range except for one ex-
treme deviate in the IECYT. The extreme deviate, experi-
ment 34, 1941 [see Federer (6), table 12], may be omitted
from the comparisons since the five double crosses in this
experiment were selected for wide differences in yielding
ability.

The frequency distributions for all three variance com-
ponents and the $\hat{a}$'s are skewed to the right. Some transfor-
mation such as the logarithmic may tend to normalize the
error variance components, but such a transformation is not
suitable for the variety and replicate variance components
and the $\hat{a}$'s due to the negative values. A logarithmic trans-
formation of the values $\hat{a}_n + 1$ might be used since the lower
limit of $\hat{a}$ is $-1/n$. 
TABLE 8. FREQUENCY DISTRIBUTION OF THE $a$’S FROM 302 CORN EXPERIMENTS INVOLVING SINGLE CROSSES, TOP CROSSES, DOUBLE CROSSES AND OTHER KINDS OF CORN.

<table>
<thead>
<tr>
<th>Class interval</th>
<th>Top cross</th>
<th>Single cross</th>
<th>Double cross</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IECYT</td>
<td>IECYT</td>
<td>IECYT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.4 - -0.201</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-0.2 - 0.001</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>0.0 - 0.199</td>
<td>8</td>
<td>4</td>
<td>19</td>
<td>16</td>
<td>47</td>
</tr>
<tr>
<td>0.2 - 0.399</td>
<td>9</td>
<td>2</td>
<td>28</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>0.4 - 0.599</td>
<td>8</td>
<td>9</td>
<td>19</td>
<td>13</td>
<td>51</td>
</tr>
<tr>
<td>0.6 - 0.799</td>
<td>5</td>
<td>10</td>
<td>11</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>0.8 - 0.999</td>
<td>3</td>
<td>13</td>
<td>3</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>1.0 - 1.199</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>1.2 - 1.399</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>1.4 - 1.599</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>1.6 - 1.799</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1.8 - 1.999</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2.0 - 2.199</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2.2 - 2.399</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class interval</th>
<th>Top cross</th>
<th>Single cross</th>
<th>Double cross</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IECYT</td>
<td>IECYT</td>
<td>IECYT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4 - 2.599</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2.6 - 2.799</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2.8 - 2.999</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3.0 - 3.199</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3.2 - 3.399</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3.4 - 3.599</td>
<td>1</td>
<td>--</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3.6 - 3.799</td>
<td>--</td>
<td>--</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3.8 - 3.899</td>
<td>--</td>
<td>--</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3.9 - 3.999</td>
<td>--</td>
<td>--</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8.9 - 9.999</td>
<td>--</td>
<td>--</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9.0 - 9.999</td>
<td>--</td>
<td>--</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15.1 - 15.299</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15.3 - 15.499</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Total number 46 73 185 70 8 302
17. Average values of the \( \hat{a} \)'s for single crosses, top crosses and double crosses.

The yearly arithmetic averages of the \( \hat{a} \)'s for the IECYT experiments vary from 0.6999 to 1.3450, around the mean \( 0.9546 \pm 0.1047 \) of the IECYT. Furthermore the mean, \( \bar{a} = 0.5212 \pm 0.0648 \) of the ICYT group of 70 experiments, is considerably lower than the mean of the IECYT experiments or of any of the yearly averages. This discrepancy may be explained, in part at least, by observing that the composition of the two groups of experiments is quite different, i.e., all of the entries in the ICYT experiments are double crosses while those in the IECYT group consist of several types of corn (see table 5). The mean of the \( \hat{a} \)'s for 46 top-cross experiments is \( 0.8180 \pm 0.1173 \), for 73 single-cross experiments \( 1.1440 \pm 0.0798 \), and for 105 double-cross experiments \( 0.6617 \pm 0.1471 \) in the IECYT group. For all 175 double-cross experiments the mean is 0.6055. The mean ratios of the IECYT and of the ICYT double crosses differ considerably. This discrepancy may be explained by the extreme deviate obtained for experiment 34 in 1941 (6); if this experiment is omitted, then the mean ratio, \( \bar{a} \), of the remaining 104 double-cross experiments in the IECYT group is \( 0.5224 \pm 0.0478 \). Thus, the means of the two groups of double crosses become almost identical if the one extreme deviate is omitted. This agreement between the two groups is not unexpected since the double crosses in the ICYT have not all been selected for high yield. It was the practice of some commercial breeders to test their double-cross combinations in the ICYT. This practice has now been restricted due to the fact that only a limited number of entries may be entered and an entry may appear in only one of the four yield test districts, that is, northern, north central, south central or southern (see section 21).

A single cross, which involves two lines only, is more specific for the expression of an inherited character, such as yield, than any of the other types of corn hybrids. It is expected that the variation among single-cross yields would be the largest for any of the corn hybrids; the data confirm this. Also, the mean of the \( \hat{a} \)'s for the top crosses falls midway between the means for single crosses and double crosses. This result also follows expectation since the top cross may be classified as a three-way cross, as it usually is defined to be the cross of an inbred line with an unrelated tester parent. The tester may be a single cross, double cross or open-pollinated variety.
The standard errors in percent of their respective means for the various types of crosses are not excessively high. The standard errors in percent of their respective means are 14.3, 7.0, 9.2, 12.6 percent for top cross, single cross, double cross from the IECYT (omitting experiment 34, 1941), and double cross from the ICYT experiments. From a practical standpoint these means, which are unbiased estimates, may be considered accurate enough to use in determining average genetic progress.

Since the mean ratios differ, the three types of crosses will be considered separately. The average genetic progress (formula 3.1) made by selecting the highest two (tables 2 and 3) single crosses, top crosses and double crosses from a sample of \( m \) varieties for varying numbers of replicates is given in table 9. For a given number of plots, the best design for single crosses apparently is one with two replicates. On the other hand, the best design for top crosses or double crosses will depend on the number of strains in the test. From table 9 the indications are that types of corn exhibiting low genetic variation will need to be measured more accur-

**TABLE 9. A COMPARISON OF POSITIVE WEIGHTED STANDARD DEVIATION UNITS FOR VARYING NUMBERS OF VARIETIES AND REPlicATES WHEN THE HIGHEST TWO YIELDING VARIETIES ARE SELECTED IN CONTRAST TO TWO RANDOMLY SELECTED ONES FROM THE SAME POPULATION (MULTIPLY EACH VALUE BY \( \hat{a}_r \)).**

<table>
<thead>
<tr>
<th>Number of varieties</th>
<th>Average value of the two largest deviates</th>
<th>Positive standard deviation units with the number of replicates indicated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>---</td>
</tr>
<tr>
<td>Single-cross experiments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.74</td>
<td>1.359</td>
</tr>
<tr>
<td>50</td>
<td>2.05</td>
<td>1.602</td>
</tr>
<tr>
<td>75</td>
<td>2.22</td>
<td>1.775</td>
</tr>
<tr>
<td>100</td>
<td>2.33</td>
<td>1.826</td>
</tr>
<tr>
<td>150</td>
<td>2.48</td>
<td>1.938</td>
</tr>
<tr>
<td>200</td>
<td>2.58</td>
<td>2.016</td>
</tr>
<tr>
<td>Top-cross experiments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.74</td>
<td>1.056</td>
</tr>
<tr>
<td>50</td>
<td>2.05</td>
<td>1.244</td>
</tr>
<tr>
<td>75</td>
<td>2.22</td>
<td>1.347</td>
</tr>
<tr>
<td>100</td>
<td>2.33</td>
<td>1.414</td>
</tr>
<tr>
<td>150</td>
<td>2.48</td>
<td>1.505</td>
</tr>
<tr>
<td>200</td>
<td>2.58</td>
<td>1.565</td>
</tr>
<tr>
<td>Double-cross experiments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.74</td>
<td>0.736</td>
</tr>
<tr>
<td>50</td>
<td>2.05</td>
<td>0.867</td>
</tr>
<tr>
<td>75</td>
<td>2.22</td>
<td>0.939</td>
</tr>
<tr>
<td>100</td>
<td>2.33</td>
<td>0.986</td>
</tr>
<tr>
<td>150</td>
<td>2.48</td>
<td>1.049</td>
</tr>
<tr>
<td>200</td>
<td>2.58</td>
<td>1.092</td>
</tr>
</tbody>
</table>
ately than will types of corn showing high genetic variation such as single crosses.

DATA FROM EXPERIMENTS INVOLVING MORE THAN ONE TESTER

18. Frequency distribution of yields.

The 11 experiments involving lines, testers and replicates [see tables 10 to 12 and Federer (6), tables 23 to 27] have been discussed by Federer and Sprague (7) with regard to the relative numbers of replicates, lines and testers.

A detailed description of each of the 11 experiments is presented by Federer (6) in table 23. The analyses given by Federer and Sprague (7), table 11, were obtained using actual field weights where the plot yields had been weighed to 1/10 pound. The field weights were rounded to whole numbers in obtaining tables 10 and 12.

In a yield trial it may be desirable to determine the normality of the data. The frequency distribution of the field weights (rounded to whole numbers) for the 11 experiments was obtained (table 10). As far as was evident from observation of the histogram of yields in any of the experiments, the frequency distributions of yields do not deviate much from normal and appear to be symmetrical in form. Occasionally yield data from corn experiments are slightly skewed, but none of these 11 experiments were.

A comparison of the frequency distributions for yields and log yields indicated that the latter tended to be skewed somewhat to the left resulting in a slightly unsymmetrical distribution.

19. Analyses of variance.

The analyses of variance and the estimates of ratios of variance components are presented in tables 11 and 12 for yields and logarithm of yields, respectively. The error variances could not be assumed to be homogeneous, since Bartlett's (1) test of homogeneity for a set of variances gave a large chi-square value for each set. Furthermore, the variation among both the line and the line by tester variance components was larger than might be attributed to chance. The test of significance used to test the variation among variance components was an adaption of Cochran's (2) Q test, where \( Q = \sum \frac{1}{s_i^2} (x_i - \bar{x})^2 \), where a simple transformation of \( Q \) may be referred to the published tables of
<table>
<thead>
<tr>
<th>Yield (lbs. per plot)</th>
<th>Randomized block design</th>
<th>Split plot design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experiment no. in 1940</td>
<td>Experiment no. in 1941</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>36</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>37</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>38</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>39</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>40</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Number</td>
<td>390</td>
<td>184</td>
</tr>
<tr>
<td>Exp. mean</td>
<td>17.83</td>
<td>20.42</td>
</tr>
</tbody>
</table>
chi-square, and where the weights are the reciprocals of the variances of the individual variance components (4). These results are not unexpected from the nature of the data, i.e., different lines and testers were compared in different locations and years, and the lines came from different kinds of parents, namely high yielding F2 populations and open-pollinated varieties.

The unbiased estimates of the ratios of variance components, \( \hat{b}, \hat{d} \) and \( \hat{e} \), as obtained from the transformed data are slightly smaller than those obtained from the original data. The slight discrepancy would not alter the comparisons among lines, testers and replicates [see Federer and Sprague (7)] to any appreciable extent. The conclusions reached would be the same whether the data were transformed or not.

The analysis in table 11 was performed under the assumption that the yield of any individual plot of a strain of corn may be expressed in a linear manner, thus

\[
y_{gh} = \mu + r_i + t_h + v_g + (vt)_{gh} + (rt)_{1h} + (rv)_{1g} + e_{1gh}, \quad (19.1)
\]

(for a randomized complete block design; the last two terms are combined for a split-plot design) where the effects are as explained previously.

Now under the assumption that the yield effects are multiplicative rather than additive, some such model as the following might be proposed:

\[
y_{gh} = (\mu + r_i + t_h + (vt)_{gh} + (rt)_{1h} + (rv)_{1g} + e_{1gh}) v_g, \quad (19.2)
\]

where \( y_g \) is the line effect of the \( g \)th line or variety. To handle such a situation as this in obtaining estimates of the various mean squares the following is proposed:

\[
\log y_{gh} = \log \{\mu + r_i + t_h + (vt)_{gh} + (rt)_{1h} + (rv)_{1g} + e_{1gh}\} + \log v_g. \quad (19.3)
\]

First remove the line or variety effect as follows: the expected value of the log yield equals the observed log yield plus the experimental mean of all log yields minus the variety mean of log yields. The next step is to obtain the antilog of the expected log yields which will give the field weights of individual plot yields adjusted for line effect. The ordinary analysis of variance on yields is then obtained except that the sum of squares among line totals and the degrees of freedom for lines will be zero.

The above type of analysis was made on experiment 3 of table 11. The tester by line mean square obtained was 4.60 as compared to 4.37 in table 11. The results do not differ greatly, and this is what would be expected from the frequency distribution of yields (table 10).
TABLE 11. DATA FROM 11 CORN TOP-CROSS EXPERIMENTS INVOLVING MORE THAN ONE TESTER.

<table>
<thead>
<tr>
<th>Year</th>
<th>Exp. No.</th>
<th>Kinds* of testers</th>
<th>Source of variation</th>
<th>Components of variance</th>
<th>Unbiased estimates of the ratio of variance components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>^σ² = error variance component</td>
<td>Line x replicate</td>
<td>Line x tester</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean square</td>
<td>d.f.</td>
<td>Mean square</td>
</tr>
<tr>
<td>1940</td>
<td>14</td>
<td>S.C.</td>
<td>4.13</td>
<td>152</td>
<td>4.08</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>S.C.</td>
<td>5.11</td>
<td>66</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>S.C.</td>
<td>3.97</td>
<td>20</td>
<td>3.96</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>S.C.</td>
<td>1.50</td>
<td>36</td>
<td>3.39</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>D.C.</td>
<td>3.82</td>
<td>138</td>
<td>4.60</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>D.C.</td>
<td>5.80</td>
<td>160</td>
<td>11.08</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>D.C.</td>
<td>2.54</td>
<td>16</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>D.C.</td>
<td>4.39</td>
<td>28</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Randomized Complete Block Designs

<table>
<thead>
<tr>
<th>Year</th>
<th>Exp. No.</th>
<th>Kinds* of testers</th>
<th>Source of variation</th>
<th>Components of variance</th>
<th>Unbiased estimates of the ratio of variance components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean square</td>
<td>d.f.</td>
<td>Mean square</td>
</tr>
<tr>
<td>1941</td>
<td>1</td>
<td>S.C.</td>
<td>9.19</td>
<td>693</td>
<td>19.16</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>S.C.</td>
<td>6.95</td>
<td>126</td>
<td>20.66</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>S.C.</td>
<td>9.84</td>
<td>370</td>
<td>15.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average of 8 experiments

<table>
<thead>
<tr>
<th>Split-plot Designs</th>
<th>^b</th>
<th>^e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1942</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Average of 3 experiments</td>
<td>0.328</td>
<td>0.615</td>
</tr>
<tr>
<td>Average of 11 experiments</td>
<td>0.229</td>
<td>0.582</td>
</tr>
</tbody>
</table>

* S.C. = Single-cross testers; D.C. = double-cross testers.
TABLE 12. DATA FROM 11 CORN TOP-CROSS EXPERIMENTS INVOLVING MORE THAN ONE TESTER (YIELDS TRANSFORMED TO LOGARITHMS AND EACH MEAN SQUARE AND VARIANCE COMPONENT MULTIPLIED BY 10^4).

<table>
<thead>
<tr>
<th>Year</th>
<th>Exp. No.</th>
<th>Kinds of testers</th>
<th>Source of variation</th>
<th>Components of variance</th>
<th>Unbiased estimates of the ratio of variance components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean square</td>
<td>d.f.</td>
<td>Mean square</td>
</tr>
<tr>
<td>1940</td>
<td>14</td>
<td>S.C.</td>
<td>29</td>
<td>152</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>S.C.</td>
<td>40</td>
<td>66</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>S.C.</td>
<td>18</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>S.C.</td>
<td>6</td>
<td>36</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>D.C.</td>
<td>10</td>
<td>138</td>
<td>11</td>
</tr>
<tr>
<td>1941</td>
<td>2</td>
<td>D.C.</td>
<td>17</td>
<td>160</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>D.C.</td>
<td>5</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>D.C.</td>
<td>56</td>
<td>28</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Average of 8 experiments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940</td>
<td>1</td>
<td>S.C.</td>
<td>22</td>
<td>693</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>S.C.</td>
<td>11</td>
<td>126</td>
<td>30</td>
</tr>
<tr>
<td>1942</td>
<td>8 &amp; 9</td>
<td>S.C.</td>
<td>22</td>
<td>970</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Average of 3 experiments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Average of 11 experiments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Another possible model to consider would be the following:

\[ Y_{i_{gh}} = (\mu + r_i + (vt)_{gh} + (rt)_{ih} + (rv)_{ig} + e_{i_{gh}}) v_h t_h. \]  

(19.4)

In this case the line and tester effect could be removed in a manner similar to the preceding case. Although the above models do not appear to fit yield data from corn hybrids, they may be satisfactory models for other types of plants or perhaps of animals.


In estimating a mean effect many methods of estimation are available. The method used by Federer and Sprague (7) and in this paper has been to obtain an arithmetic mean of the unbiased estimates of a ratio. To determine the reliability of this estimate it is necessary to compare it with an efficient estimate such as the maximum likelihood estimate when it exists and is suitable for the data under consideration.

For the data of table 11 the following estimate of the mean ratio is assumed to be appropriate:

\[ \bar{b'} = \frac{1}{N} \sum_{u=1}^{N} \hat{b}_u. \]  

(20.1)

Under the assumption that the \( \hat{b}_u \) are all estimates of the same parameter \( \beta \), the variance of \( \bar{b'} \) is obtained from formula (2.3), thus

\[ V(\bar{b'}) = \frac{1}{N^2} \sum_{u=1}^{N} V(\hat{b}_u) \]

\[ = \frac{2}{N^2} \sum_{u=1}^{N} \left[ \frac{f_\theta + f_{i_1t} - 2}{f_{i_1t} f_\theta - 4f_{i_1t}} \right] u \left[ \frac{1}{n_u} + \hat{b}_u \right]^2. \]

(20.2)

If the \( \hat{b}_u \) are assumed to be estimates of the \( \beta_u \) and the \( \beta_u \) are not all equal, then the variance of \( \bar{b} = \frac{1}{N} \sum \hat{b}_u \) is

\[ V(\bar{b}) = \sum_{u=1}^{N} \frac{V(\hat{b}_u)}{N(N-1)}. \]  

(20.3)

*The reader is referred to J. W. Tukey, Biometrics 5:232, 1949, for a more recent view on the subject of non-additivity.
The ratio \( V(\widehat{b})/V(\widehat{b}') \) is suitable as a test of significance of the difference of the ratios if \( f_{it} \) and \( f_c \) are not too small, say less than 20 to 30. The resulting \( F \) ratio will have \( N-1 \) degrees of freedom associated with the numerator and infinite degrees of freedom associated with the denominator, \( V(\widehat{b}') \). The four examples in table 13 fit this criterion but the one in table 11 does not.

Biological theory (7) indicates that the ratios in table 11 should be different. Bartlett's (1) chi-square test of \( \sigma^2 \)'s resulted in a large value of chi-square. A chi-square test on the \( \sigma^2 \)'s indicated that these variance components were unusually discrepant. Since the numerators in the ratios were different and since the denominators were different, the ratios must be different. Therefore \( \hat{b}_u \) was assumed to be an estimate of \( \beta_u \) and all \( \beta_u \) are not assumed to be equal. The test suggested in the preceding paragraph does not agree with the above for the data of table 11. However, when the four experiments with \( f_{it} < 20 \) are omitted (example 1, table 13), both tests agree.

Several other estimates of the mean are available. For example, if all \( \beta_u \) are assumed equal, a maximum likelihood estimate, \( \hat{\beta} \), of \( \beta \) may be obtained from the equation

\[
\frac{N}{\sum_{u=1}^{N} \left( \frac{(nf_e)_u}{1 + n_u \beta} - \frac{n_u (f_e + f_{1t})}{1 + n_u \beta + \left( \frac{f_{1t}}{f_e} \right)_u (1 + n_u \beta)} \right)} = 0. \tag{20.4}
\]

The estimate has a variance

\[
V(\hat{\beta}) = \frac{N}{\sum_{u=1}^{N} \left[ \frac{n_u \hat{\beta}}{1 + n_u \beta} \right]^2 \left[ \frac{f_e f_{1t}}{f_e + f_{1t} + 2} \right]_u}. \tag{20.5}
\]

Other estimates that might be considered are

\[
\overline{b}_w = \frac{\Sigma (f_e \hat{b})_u}{\Sigma (f_e)_u}, \tag{20.6}
\]

\[
\overline{b}_w' = \frac{\Sigma (f_{1t} \hat{b})_u}{\Sigma (f_{1t})_u}, \tag{20.7}
\]

and

\[
\overline{b}_w* = \frac{\Sigma (f_e + f_{1t})_u \hat{b}_u}{\Sigma (f_e + f_{1t})_u}. \tag{20.8}
\]
The variance of the above three estimates would be of the form

$$\frac{\sum w_i (\hat{b}_n - \bar{b}_n)^2}{N \sum w_i}$$

where \( w_i \) represents the weights appropriate for each estimate. For the data of table 11, \( \bar{b}_w' \) was found to be .227, with a variance \( V(\bar{b}_w') = .00225 \). This variance is lower than the unweighted variance obtained for these data (see table 13).

A first approximation of an estimate of \( \hat{\beta} \) in equation (20.4) may be obtained from either equation (20.6) or equation (20.8) although the latter was closer to \( \hat{\beta} \) in the four examples of table 13; the values of \( \bar{b}_w^* \) were 0.199, 0.685, 0.1046 and 2.139 for the four examples of table 13 and 0.200 for the data of table 11.

Still another type of estimate of the mean ratio of variance components is available from formula (2.4). For the data of table 11, \( \bar{b}'' = \bar{b}_u^2 / \bar{b}_e^2 = 1.2694/5.2036 = 0.244 \). The variance of \( \bar{b}'' \) may be approximated from formula (2.5). It was found to be .004146, which is about 1.6 times \( V(\bar{b}) = .00257 \). Similarly \( \bar{d}'' \) was obtained as 0.536 with a variance equal to 0.041031, which is about 1.4 times larger than \( V(\bar{d}) = 0.028614 \).

Now it appears that \( V(\bar{b}') < V(\bar{b}) \) unless the individual \( \hat{b}_n \) are nearly all equal as in example 3 or if the \( f_{it} \) are small. In the former case \( V(\bar{b}) \) could equal zero while \( V(\bar{b}') \) could be a value quite different from zero. Likewise \( V(\hat{\beta}) \) lies between \( V(\bar{b}') \) and \( V(\bar{b}) \) except in the last example, where divergent values such as \( \hat{b}_n = 15.442 \) cause undue effect on \( V(\bar{b}') \) and \( V(\bar{b}) \) while they have little effect on \( V(\hat{\beta}) \).

It may be instructive to draw random samples from the 302 individual ratios presented by Federer (6). After drawing a large number of samples and computing the various estimates and their variances, it may be possible to generalize comparisons involving the methods of estimation. One important point observable from table 13 is that the variance may be considerably underestimated if the wrong variance formula is chosen. Further investigation of this problem would be desirable.

For the data reported here, formula (20.1) is considered to give an estimate of the mean ratio, and formula (20.3) is
### TABLE 13. EXAMPLES SELECTED TO SHOW RELATIONSHIPS OF ESTIMATES OF MEAN RATIOS AND THEIR VARIANCES.

<table>
<thead>
<tr>
<th>Example 1.* Experiments from table 11 for which ( f_v &gt; 20 )</th>
<th>Example 2. First 10 experiments in the 1940 IEICYT group [table 11 of thesis (6)]</th>
<th>Example 3. Selected uniform values of the ratios ( \hat{a} = \hat{b} ) [table 28 of thesis (6)]</th>
<th>Example 4. Selected divergent values of the ratios ( \hat{a} ) from the 1945 IEICYT group [table 16 of thesis (6)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp.**</td>
<td>Exp. no.</td>
<td>( n )</td>
<td>( f_v = f_{i1} )</td>
</tr>
<tr>
<td>'40-14</td>
<td>5</td>
<td>152</td>
<td>38</td>
</tr>
<tr>
<td>'40-15</td>
<td>4</td>
<td>66</td>
<td>22</td>
</tr>
<tr>
<td>'41-1</td>
<td>3</td>
<td>138</td>
<td>69</td>
</tr>
<tr>
<td>'41-2</td>
<td>3</td>
<td>150</td>
<td>80</td>
</tr>
<tr>
<td>'40-1</td>
<td>4</td>
<td>693</td>
<td>154</td>
</tr>
<tr>
<td>'40-45</td>
<td>3</td>
<td>126</td>
<td>42</td>
</tr>
<tr>
<td>'41-2</td>
<td>3</td>
<td>150</td>
<td>80</td>
</tr>
<tr>
<td>'40-1</td>
<td>4</td>
<td>693</td>
<td>154</td>
</tr>
<tr>
<td>'40-45</td>
<td>3</td>
<td>126</td>
<td>42</td>
</tr>
<tr>
<td>'42-8&amp;9</td>
<td>6</td>
<td>970</td>
<td>97</td>
</tr>
</tbody>
</table>

Average = \( \overline{a} = \overline{b} = 0.242 \):

Max. likelihood est. =

\[
\hat{a} = \hat{b} = 0.195
\]

Aver. wt'd. by \( f_v \) = 0.192

\[
V(\overline{a}') = V(\overline{b}') = .00250
\]

\[
V(\overline{a}) = V(\overline{b}) = .00599
\]

\[
V(\hat{a}) = V(\hat{b}) = .00341
\]

\[
\text{Average } = a = b = 0.242
\]

\[
\text{Max. likelihood est. } = a = b = 0.195
\]

\[
\text{Aver. wt'd. by } f_v = 0.192
\]

\[
V(\overline{a}') = V(\overline{b}') = .00250
\]

\[
V(\overline{a}) = V(\overline{b}) = .00599
\]

\[
V(\hat{a}) = V(\hat{b}) = .00341
\]

*For all 11 experiments \( a = 0.229, \hat{a} = 0.198, V(\overline{a}') = .00385, V(\overline{a}) = .00257, \) and \( V(\hat{a}) = .00516. \) The average of the \( \hat{a} \) weighted by \( f_v \) is 0.193.

**First number refers to year and the second to experiment number.
considered to give an estimate of the variance of the mean. However, it was deemed instructive to compare the various estimates and their variances on some experimental results. This was done in table 13.

DATA FROM EXPERIMENTS INVOLVING MORE THAN ONE PLACE OR LOCATION

21. Description of the data.

For purposes of conducting yield trials, Iowa is divided into four sections with three districts to a section. Districts 1, 2 and 3 make up the northern section of the state; districts 4, 5 and 6 the north central section; districts 7, 8 and 9 the south central section; and districts 10, 11 and 12 the southern section of Iowa. As a result of the performance of the entries in the Iowa Corn Yield Tests, corn hybrids or varieties are recommended for each of the four sections of Iowa; the sections are of approximately equal area.

During the years 1940 through 1946, six to nine replicates were planted for each experiment in the Iowa Corn Yield Tests [see Federer (6), table 18]. The entries were tested in all three districts of a section. Thus for a basis of comparison, three places or locations with six replicates at a place will be used as the standard in the following discussion. The design was considered to be the randomized complete block even though lattice designs were used exclusively in the Iowa Corn Yield Tests. Cochran (3) has compared the accuracy of lattice designs for corn yield trials in Iowa.

Over the period of years 1940 through 1946, 38 of the IECYT experiments [see Federer (6), table 28] were planted at more than one place or location. There were no missing plots in any of the 38 experiments at any of the locations. The analysis of these data is straightforward, and the expectation of the mean squares is the same as that given for design III of table 1. The analysis of data with missing plots at one or more of the places is not straightforward and has not been worked out. Therefore none of the ICYT experiments and several of the IECYT experiments involving more than one location could not be used for determining the relative magnitudes of the error to the variety by place and variety variance components.

The data on the 38 IECYT experiments which were planted at more than one location and had no missing yields have been given by Federer (6) in table 28. From these data the calculation of the estimated variance components, $\hat{\sigma}_n^2$ and $\hat{\sigma}_v^2$, and of the means of the ratios $\hat{b}$ and $\hat{d}$ [formulas (1.9)}
and \((1.11)\) was straightforward. The variance of the mean ratios \(\bar{b}\) was approximated by the formula \(\frac{\sum_{u=1}^{N}(\hat{b}_u - \bar{b})^2}{N(N-1)}\).

The variance of \(\bar{d}\) was computed from a similar formula.

### 22. Relative value of combinations of strains, replicates and locations in corn yield trials.

The means \(\bar{b} = 0.2502 \pm 0.0503\) and \(\bar{d} = 0.4197 \pm 0.0492\), obtained from the summarization of the data from the 38 IECYT experiments, may be expected to be applicable to various types of corn yield tests conducted under similar conditions. The results have broader implications than might first be realized. As stated in section 21, Iowa is divided into four areas with respect to corn yield tests. It might be argued that the division of the four areas is not the one yielding the most homogenous results. For example, the northern section has a shorter growing season than the remaining three sections, but within the section climatic conditions vary from west to east. It could be argued that districts 1, 2 and 3 should not be put into the same section for yield purposes. These arguments are sound if the experiments were to be conducted year after year. This is not the case, and the corn breeders have partially insured themselves against the year-to-year variation by selecting districts which varied with regard to climatic factors. Even though the state may be rezoned for other purposes, it is suggested that the present division of the state be retained for corn yield test purposes unless the experiments are to be repeated year after year.

The means \(\bar{b}\) and \(\bar{d}\) were computed from all the 38 experiments and were used to obtain tables 14 and 15. Since the composition of the 38 experiments varied with regard to kind of cross, it might be argued that the results should be treated separately for single crosses, top crosses and double crosses, for which \(\bar{b} = 0.3820, 0.4112\) and \(0.1323\), respectively. This was not done because of the small number of experiments involved for each type of hybrid. The mean ratios \(\bar{d} = 0.7138, 0.3226\) and \(0.3307\), respectively, for single crosses, top crosses and double crosses should be treated separately if enough observations on each type of cross were available. It would be expected that \(\bar{d}\) from top-cross experiments would be intermediate to the other two values of \(\bar{d}\). In any
### Table 14. A Comparison of Positive Weighted Standard Deviation Units for Varying Numbers of Varieties, Locations and Replicates When the Highest Yielding Variety is Selected in Contrast to a Randomly Selected One from the Same Population. (Multiply each value by $\hat{\sigma}$.)

<table>
<thead>
<tr>
<th>Number of varieties</th>
<th>Average value of largest deviate</th>
<th>Positive standard deviation units with the number of locations or places indicated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1 replicate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.97</td>
<td>0.640</td>
</tr>
<tr>
<td>50</td>
<td>2.25</td>
<td>0.731</td>
</tr>
<tr>
<td>75</td>
<td>2.40</td>
<td>0.780</td>
</tr>
<tr>
<td>100</td>
<td>2.51</td>
<td>0.815</td>
</tr>
<tr>
<td>150</td>
<td>2.65</td>
<td>0.861</td>
</tr>
<tr>
<td>200</td>
<td>2.75</td>
<td>0.893</td>
</tr>
<tr>
<td>300</td>
<td>2.90</td>
<td>0.942</td>
</tr>
<tr>
<td>500</td>
<td>3.05</td>
<td>0.991</td>
</tr>
<tr>
<td>2 replicates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.97</td>
<td>0.764</td>
</tr>
<tr>
<td>50</td>
<td>2.25</td>
<td>0.873</td>
</tr>
<tr>
<td>75</td>
<td>2.40</td>
<td>0.931</td>
</tr>
<tr>
<td>100</td>
<td>2.51</td>
<td>0.974</td>
</tr>
<tr>
<td>150</td>
<td>2.65</td>
<td>1.028</td>
</tr>
<tr>
<td>200</td>
<td>2.75</td>
<td>1.067</td>
</tr>
<tr>
<td>300</td>
<td>2.90</td>
<td>1.125</td>
</tr>
<tr>
<td>500</td>
<td>3.05</td>
<td>1.183</td>
</tr>
<tr>
<td>3 replicates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.97</td>
<td>0.825</td>
</tr>
<tr>
<td>50</td>
<td>2.25</td>
<td>0.943</td>
</tr>
<tr>
<td>75</td>
<td>2.40</td>
<td>1.066</td>
</tr>
<tr>
<td>100</td>
<td>2.51</td>
<td>1.082</td>
</tr>
<tr>
<td>150</td>
<td>2.65</td>
<td>1.110</td>
</tr>
<tr>
<td>200</td>
<td>2.75</td>
<td>1.152</td>
</tr>
<tr>
<td>300</td>
<td>2.90</td>
<td>1.218</td>
</tr>
<tr>
<td>500</td>
<td>3.05</td>
<td>1.278</td>
</tr>
<tr>
<td>6 replicates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.97</td>
<td>0.904</td>
</tr>
<tr>
<td>50</td>
<td>2.25</td>
<td>1.032</td>
</tr>
<tr>
<td>75</td>
<td>2.40</td>
<td>1.101</td>
</tr>
<tr>
<td>100</td>
<td>2.51</td>
<td>1.152</td>
</tr>
<tr>
<td>150</td>
<td>2.65</td>
<td>1.216</td>
</tr>
<tr>
<td>200</td>
<td>2.75</td>
<td>1.262</td>
</tr>
<tr>
<td>300</td>
<td>2.90</td>
<td>1.331</td>
</tr>
<tr>
<td>500</td>
<td>3.05</td>
<td>1.392</td>
</tr>
</tbody>
</table>
event, these data illustrate the procedures for the summarization of a group of experiments.

In determining the procedure to follow in obtaining the highest yielding hybrids on the average, the formulas developed in sections 3 and 4 are applicable to experiments involving more than one location. If the highest yielding variety or hybrid is selected rather than a randomly selected one from a group of $m$ varieties, the average genetic advance or progress or the average advance in the selection of high combining or yielding ability may be computed from formula (3.2). If one wishes to select the highest two, then take the average value of the highest two deviates (tables 2 and 3) for the value of $x_m$.

Table 14 has been computed using formula (3.2) for average genetic advance. The comparison of the relative number of replicates and locations is the one of interest. For example, greater progress in selecting high yielding hybrids may be made by using six places and two replicates at a place rather than the standard of three places with six replicates at a place. The former scheme involves only two-thirds as many plots as the latter arrangement. Without considering the relative costs of more locations and fewer replicates, an experimenter would make better use of his efforts by using six locations and two replicates at a place rather than three locations and six replicates at a place.

The desirability of using more locations is further stressed by the fact that the yield results in one or more districts frequently are lost. For example, in district 11, which is the central district in the southern section of Iowa, no yield results were obtained in six of the last eight years [see Federer (6), table 18; no results were obtained in 1947, either]. It would be highly desirable, then, to use more districts in the southern section of Iowa since only two locations were used in the majority of years. Regardless of additional cost for planting more trials in the southern section, it appears that more places are necessary in order to obtain yield estimates for this area.

23. Relative value of combinations of replicates and locations in corn yield tests.

In obtaining an idea of the relative efficiency of a design with three places and six replicates at a place to various other designs involving replicates and places, table 15 was prepared. The relative efficiencies were obtained from the comparison of the error variances of a variety mean. Since the relative efficiencies represent a ratio, the error variance
of a variety mean may be expressed as

\[ 2 \left\{ \frac{\hat{\sigma}_e^2}{np} + \frac{\hat{\sigma}_v^2}{p} \right\} = 2\sigma^2 \left\{ \frac{1}{np} + \frac{0.2502}{p} \right\} \]

where the symbols are as defined previously and 0.2502 equals \( \bar{b} \) (see section 22). Thus by varying \( n \) and \( p \) various error variances may be obtained. The efficiencies in table 15 are given in percent of the error variance for a variety mean from three locations with six replicates at a place, which is \( 2\sigma^2 (0.1390) \). The relative efficiency of using six places with two replicates at a place to the standard is

\[ \left\{ \frac{1}{18} + \frac{0.2502}{3} \right\} \left(\frac{1}{12} + \frac{0.2502}{6}\right) (100) = \left(\frac{0.139}{0.125}\right) (100) = 111 \text{ percent}. \]

Table 15 presents several interesting features. Nine locations with one replicate at a location is as efficient as the standard three places with six replicates at a place. Furthermore, 24 replicates at one place is about one-half as efficient as the standard. To make this comparison even more striking, if an infinity of replicates were used at one place, the standard still would be the better plan to use. Since the experimenter usually would desire more than one replicate at a place, the arrangement of six locations with two replicates at a place would appear to be the preferable design. Although only two-thirds as many plots are planted and harvested, the additional cost of more locations may offset the apparent increase in efficiency. Further studies of the year effect on a variety yield and costs of planting, harvesting and travel need to be conducted in order to make concrete suggestions to be put into practice.

**Table 15. Relative Efficiency (in Percent) of a Randomized Complete Block Design at Three Places with Six Replicates at a Place to Other Combinations of Replicates and Places.**

<table>
<thead>
<tr>
<th>Number of places</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
</tr>
<tr>
<td>8</td>
<td>89</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
</tr>
</tbody>
</table>
LITERATURE CITED


