1948

Estimation of general, specific and maternal combining abilities in crosses among inbred lines of swine

Charles Roy Henderson

Iowa State College

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

Part of the Agriculture Commons, Animal Sciences Commons, and the Genetics Commons

Recommended Citation


http://lib.dr.iastate.edu/rtd/12780

This Dissertation is brought to you for free and open access by Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
NOTE TO USERS

This reproduction is the best copy available.

UMI®
ESTIMATION OF GENERAL, SPECIFIC AND MATERNAL COMBINING ABILITIES IN CROSSES AMONG INBRED LINES OF SWINE

by

Charles Roy Henderson

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major Subjects: Animal Breeding Genetica

Approved:

In Charge of Major Work

Heads of Major Departments

Dean of Graduate College

Iowa State College

1948
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform DP11842
Copyright 2005 by ProQuest Information and Learning Company. All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346
# TABLE OF CONTENTS

## I. INTRODUCTION

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

## II. STATISTICAL METHODS

### A. The Mathematical Model

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
</tr>
</tbody>
</table>

### B. Statistical Problems in the Analysis of the Single Cross Data

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

### C. The Least Squares Analysis of Non-Orthogonal Data

| 1. Principles of least squares and the normal equations | 15 |
| 2. Estimation of parameters and tests of hypotheses | 16 |
| 3. Computational shortcuts in the solution of least squares equations | 21 |
| 4. Sampling errors of least squares estimates | 27 |

### D. Adjustment of Least Squares Estimates When Effects are Randomly Drawn from Populations with Known Variances

| 1. The maximum likelihood solution | 35 |
| 2. Computational procedure for adjusting least squares estimates | 38 |

### E. Estimation of a Parameter Assumed to be a Linear Function of Randomly Drawn Effects

| 39 |

### F. Estimation of Components of Variance in the Non-Orthogonal Case

| 1. Estimates from a conventional analysis of variance | 42 |
| 2. Estimates from least squares sums of squares | 44 |

### G. Application of the Methods to the Analysis of Single Cross Data

| 49 |
| 1. Tests of hypotheses | 50 |
| 2. Least squares estimates | 54 |
| 3. Sampling errors of the least squares estimates | 57 |
| 4. Adjustment of the least squares estimates | 61 |
| 5. Estimation of components of variance | 64 |
| 6. Prediction of line cross performance | 68 |
III. ANALYSIS OF EXPERIMENTAL RESULTS

A. Source and Scope of the Data .................................... 102

B. Methods of Analysis Pertaining to the Available Single Cross Data ........................................ 105

C. Tests of Significance .................................................. 112

1. Test of hypothesis that $r_{ij} = 0$ ................................ 113
2. Test of hypothesis that $s_{ij} = 0$ ................................ 114
3. Test of hypothesis that $m_{j} = 0$ ................................ 115
4. Test of hypothesis that $g_{i} = 0$ ................................ 116

D. Least Squares Estimates ............................................. 117

E. Sampling Errors of the Least Squares Estimates ................. 124

F. Estimates of Variances .............................................. 126

G. Adjustment of Least Squares Estimates ............................ 132

IV. DISCUSSION AND APPLICATIONS .................................... 136

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. Interpretation of Experimental Results</td>
<td>138</td>
</tr>
<tr>
<td>1. General combining ability</td>
<td>140</td>
</tr>
<tr>
<td>2. Maternal effects</td>
<td>143</td>
</tr>
<tr>
<td>3. Specific effects</td>
<td>145</td>
</tr>
<tr>
<td>4. Sex linkage effects</td>
<td>146</td>
</tr>
<tr>
<td>C. Utilization of the Estimates of Line and Cross Line Characteristics</td>
<td>146</td>
</tr>
<tr>
<td>1. Selection of lines for top crossing</td>
<td>146</td>
</tr>
<tr>
<td>2. Selection of lines for use as female parents</td>
<td>147</td>
</tr>
<tr>
<td>3. Selection of specific crosses</td>
<td>148</td>
</tr>
<tr>
<td>D. Choice of an Efficient Testing Program</td>
<td>149</td>
</tr>
<tr>
<td>1. Value of further testing of existing lines</td>
<td>150</td>
</tr>
<tr>
<td>2. Determining the optimum number of lines to test</td>
<td>152</td>
</tr>
<tr>
<td>3. Top cross vs. single cross test</td>
<td>154</td>
</tr>
<tr>
<td>4. Early culling of lines followed by further testing</td>
<td>155</td>
</tr>
<tr>
<td>E. Further Applications of the Method of Selecting Simultaneously for Both General and Specific Effects</td>
<td>156</td>
</tr>
<tr>
<td>V. SUMMARY</td>
<td>159</td>
</tr>
<tr>
<td>VI. LITERATURE CITED</td>
<td>163</td>
</tr>
<tr>
<td>VII. ACKNOWLEDGEMENT</td>
<td>166</td>
</tr>
<tr>
<td>APPENDIX I. NUMERICAL EXAMPLE OF ANALYSIS OF SINGLE CROSSES</td>
<td>167</td>
</tr>
<tr>
<td>A. Tests of Hypotheses</td>
<td>168</td>
</tr>
<tr>
<td>1. Test of hypothesis that $r_{ij} = 0$</td>
<td>168</td>
</tr>
<tr>
<td>2. Test of hypothesis that $s_{ij} = 0$ assuming $r_{ij} = 0$</td>
<td>174</td>
</tr>
<tr>
<td>3. Test of hypothesis that $s_{ij} = 0$ assuming $m_{ij} = r_{ij} = 0$</td>
<td>175</td>
</tr>
<tr>
<td>4. Test of hypothesis that $m_{ij} = 0$ assuming $r_{ij} = 0$</td>
<td>177</td>
</tr>
<tr>
<td>5. Test of hypothesis that $m_{ij} = 0$ assuming $s_{ij} = r_{ij} = 0$</td>
<td>177</td>
</tr>
<tr>
<td>6. Test of hypothesis that $s_{ij} = 0$ assuming $r_{ij} = 0$</td>
<td>178</td>
</tr>
<tr>
<td>7. Test of hypothesis that $s_{ij} = 0$ assuming $r_{ij} = s_{ij} = m_{ij} = 0$</td>
<td>178</td>
</tr>
</tbody>
</table>
B. Least Squares Estimates .................................................. 178
1. Estimates of $g_1$, $m_j$, $s_{ij}$, $r_{ij}$ ........................................ 178
2. Estimates of $g_1$, $m_j$, $s_{ij}$ assuming $r_{ij} = 0$ .............. 181
3. Estimates of $p$, $g_1$, $m_j$ assuming $s_{ij} = r_{ij} = 0$ ....... 182

C. Sampling Errors ............................................................ 183

D. Variance Components ..................................................... 186
1. Estimation of $\sigma_e^2$ ..................................................... 187
2. Estimation of $\sigma_s^2$ ..................................................... 187
3. Estimation of $\sigma_m^2$ ..................................................... 188
4. Estimation of $\sigma_g^2$ ..................................................... 189

APPENDIX II. NUMERICAL EXAMPLE OF ANALYSIS OF BALANCED
SINGLE CROSS DESIGN ......................................................... 190

A. Least Squares Estimates .................................................. 190

B. Tests of Hypotheses ....................................................... 192
1. Test of the hypothesis that $r_{ij} = 0$ assuming $m_j \neq 0$ .................. 192
2. Test of the hypothesis that $s_{ij} = 0$ assuming $m_j \neq 0$ .................. 193
3. Test of the hypothesis that $m_j = 0$ .................................. 193
4. Test of the hypothesis that $g_1 = 0$ assuming that $m_j \neq 0$ .................. 194

C. Estimation of Variance Components ................................. 194

D. Estimation of $\sigma_{gm}$ ................................................. 195

E. Construction of Indexes for Selecting Lines and Crosses .......... 195
I. INTRODUCTION

The successful employment of inbreeding by the founders of our present day livestock breeds and the commercial success of hybrid corn have led to some intensive investigations into the possibilities of utilizing these methods to improve swine. Three important problems have arisen in connection with these studies. (Hazel and Lush, 1948.)

1. How best to make inbred lines. This problem is closely related to the question of the effects of inbreeding.

2. How best to test inbred lines. For example, can lines be judged on their own performance or must they be tested in matings with other lines or with non-inbreds? If own performance is an inadequate test, what type of tester stock should be used?

3. How best to utilize inbred lines in commercial pork production.

The effects of inbreeding have been investigated very extensively by several of the experiment stations of the North Central States in cooperation with the United States Department of Agriculture. The data from these experiments demonstrate that inbreeding is accompanied by marked loss of sow productivity. The number of pigs weaned per litter has decreased linearly with increase in coefficient of inbreeding. (Dickerson et al., 1947.) The effect of inbreeding on the weight of the individual pig has been less pronounced than the effect on litter number. It appears probable from these studies that the most efficient way to make inbred lines is to make many small lines and to inbreed rapidly, rather than to make fewer large
lines with slower rates of inbreeding. (Hazel and Lush, 1948.)

The problem of how best to test lines has received considerably less attention than has the problem of how to make the lines. Some of the theoretical aspects of the problem of testing were studied by Dickerson (1942). He calculated the numbers of animals needed to obtain statistically significant differences between two inbred lines when such lines are compared in line crosses and top crosses. Dickerson concluded that for traits like weight and productivity, which are subject to large environmental variations, the homozygosity of the lines and the kind of testing program have little influence on the number of animals required for significance of a given difference. He also concluded that little is to be gained by advancing the inbreeding much beyond 30-40 per cent before testing the lines, the exact level of inbreeding depending upon the kind of tests to be made, the significance level chosen, and the importance of dominance and epistasis.

Dickerson, Lush, and Culbertson (1946) and Winters et al (1944) have reported results of crossing inbred lines of swine. These studies were designed to compare the performance of line crosses with the performance of inbreds and non-inbreds rather than to appraise individual lines or specific crosses. The studies on crosses among inbred lines of guinea pigs reported by Wright (1922) and by Eaton (1941) were also designed to compare line crosses with inbreds.

Sprague and Tatum (1942) have estimated the variances due to general combining ability and specific combining ability in single crosses among inbred lines of corn. They found that in unselected lines the general combining ability variance component was about equal in size to the specific
combining ability component. In contrast, among lines which had previously been selected on the basis of their general combining ability, the specific combining ability component was larger than the general combining ability component. Federer and Sprague (1947) have investigated the efficiency of different designs with respect to number of testers, replications, and lines for selecting inbred lines of corn with the highest general combining ability. Apparently the design of testing programs to select simultaneously for general combining ability and specific combining ability has not been investigated. This is the type of selection involved if the goal is the development of specific crosses for commercial utilization.

The swine breeder could make better decisions with respect to the number of lines to select and the type of testing program to conduct if the following questions could be answered:

1. What use can be made of information concerning the performance of inbred lines? How valuable is such data in predicting the value of the lines in top crosses and in crosses with other lines?

2. How accurately can the value of specific single crosses and three-way crosses be predicted from knowledge of the lines' own performances and from knowledge of the performances of the lines in top crosses?

3. How useful are data from single cross tests for predicting the performance of the lines in top crosses and for predicting the outcome of particular three-way crosses?

4. Assuming that the way in which inbred lines are to be used in commercial pork production has been decided, how many
lines should be made, when should they be tested, what kind of tests or combination of tests should be designed, and how can the maximum amount of information be obtained from the results of the tests?

The answers to these questions would seem to require (1) a large number of inbred lines available for testing, (2) extensive data on the own performance of these lines, (3) extensive top cross, single cross, and three-way cross tests of the lines, (4) estimation of line and cross line parameters in each of the above tests, (5) estimation of correlations between parameters estimated in one type of test and corresponding parameters estimated in the other types of tests, (6) estimation of the sizes of the several sources of variability in the data from different types of tests, (7) construction of an index which correctly utilizes the above estimates.

Data are being accumulated at the Iowa Agricultural Experiment Station on own performance, top cross performance, single cross performance, and three-way cross performance for 12 different inbred lines. More than 200 single cross litters and a somewhat smaller number of three-way cross litters have already been produced and tested at the station, and more than 400 top cross litters have already been tested on Iowa farms. Expansion of single and three-way cross tests is planned. Consequently, it shall be possible during the next few years to obtain some reasonably good estimates of the usefulness of different types of tests for predicting the performance of lines used in different ways.

It was decided when the present study of line testing was undertaken that the best plan to start was an analysis of the existing single cross
data. There were at that time more data available on single cross tests than on three-way or top cross tests. Furthermore, it was thought that methods of analysis which might be developed for single crosses would be applicable with slight modifications to later analyses of three-way and top cross data. Still another factor determining the choice of the single cross test was that this type of test seemed to offer the best opportunities for estimating maternal and specific effects and for determining the correlation between genic values and maternal values of inbred lines.

The studies on single crosses to be reported here were designed to answer, if possible, the following questions:

1. What are the best estimates of the additive genetic values of the 12 inbred Poland China lines developed at the Iowa Agricultural Experiment Station?

2. What are the best estimates of the maternal abilities of the lines?

3. Is there any evidence that specific effects are present in crosses among the 12 inbred lines? That is, are the progeny of specific crosses any different from what would be expected on the basis of differences in general combining ability, maternal abilities, sex linkage effects, and sampling errors?

4. Is there any evidence that sex linkage increases the variability among specific crosses?

5. What are the relative sizes of the variances due to differences among lines in general combining ability and maternal ability, to differences among single crosses in specific
combining abilities and sex linkage effects, and to error?

6. What are the most probable values of potential top cross progeny of the different lines and the most probable values of potential specific single crosses among the lines?

7. What is the best testing program for estimating general combining abilities of inbred lines?

8. What is the most efficient testing program for maximizing the probability of obtaining the best specific cross when selection is based on the results of single cross tests?

It was possible to obtain answers to each of the above questions. In so doing it was necessary to set up a mathematical model which appeared to fit the biology of the material reasonably well, to utilize least squares theory to estimate parameters and to test various hypotheses, to estimate components of variance from multiple classifications with disproportionate subclass numbers, and to extend the theory of constructing selection indexes to include selection for specific (interaction) effects and to make use of complex least squares estimates as the dependent variables in these selection indexes.

The computational procedures appropriate for the analysis of non-orthogonal data are very laborious. Since the data from most breeding experiments with large animals are likely to be unbalanced, considerable attention was given to developing computational short cuts. These have been presented in sufficient detail so they may be used by other workers interested in similar problems. Computational procedures for the analysis of balanced tests of inbred lines have also been presented for the benefit of
other workers who are able to utilize balanced designs for line testing. The saving in labor in the analysis of balanced single cross and three-way designs as compared to unbalanced designs is so great that it might be advisable in some situations to discard enough data to balance the design.
II. STATISTICAL METHODS

A. The Mathematical Model

In order to obtain answers to the eight questions proposed in Section I it is necessary to set up a mathematical model which appears to be a reasonable description of the underlying biology and which is amenable to statistical treatment. The following linear hypothesis was therefore assumed.

\[ y_{ijkl} = \mu + \xi_i + \eta_j + \xi_j + \tau_{ij} + \alpha_k + \epsilon_{ijkl} \]

\( y_{ijkl} \) denotes the value of the \( i^{th} \) experimental unit of the progeny of a mating between a male of the \( i^{th} \) line by a female of the \( j^{th} \) line and falling in the \( k^{th} \) class. The experimental unit is either a litter or an individual. The value of the experimental unit can be expressed by such measurements as litter number at various ages, litter weight at various ages, individual weight at different ages, or any of these measurements expressed as logarithms or other transformations of the measurements.

It is assumed for purposes of this study that we are dealing with some hypothetical population of inbred lines from which the twelve inbred Poland China lines of the Iowa Agricultural Experiment Station have been randomly drawn. For example, it could be assumed that these twelve lines are members of a population of lines which might have been formed using as a foundation the better pure bred Poland China swine in the corn belt during the period 1930 to 1935. The advantage of making this assumption is that it is thereby possible to generalize on the results. If, in contrast, it were assumed that the twelve lines constituted an entire population, the results could apply only to the particular lines studied.

\( \mu \) is an effect common to all litters. If the number of lines is large, all of the lines of the hypothetical population of lines are mated in all
possible crosses, and the number of progeny is large and the same for each of the possible crosses, the mean of all such crosses approaches \( \mu \).

\( g_i(g_j) \) is an effect common to all progeny of the \( i^{th} \) \((j^{th}) \) line. \( g_i \) can be assumed to be one-half the additive genetic value (breeding value) of the \( i^{th} \) line, the value being expressed as the deviation from \( \mu \). If the number of lines is large, all possible crosses are made among lines, and the number of progeny per cross is equal and large, the mean of the progeny of the \( i^{th} \) line of males approaches \( \mu + g_i \). If the \( m_j = 0 \), the mean of the progeny of the \( i^{th} \) line of dam also approaches \( g_i \). It is assumed that the \( g_i \) are normally distributed with mean = 0 and variance = \( \sigma_g^2 \). Consequently, the lines of this study having been assumed to be randomly drawn, \( \mathbb{E}g_ig_j = 0 \),

where \( \mathbb{E} \) denotes expected value. Since \( g_i \) = one-half the additive genetic value of the \( i^{th} \) line, \( \sigma_g^2 \) = one-fourth of the additive genetic variance among lines. If it is assumed that there has been no selection either within or among lines, the additive genetic variance among lines is expected to be \( 2f \sigma_g^2 \), where \( f \) = the inbreeding of the progeny and \( \sigma_g^2 \) = the genetic variance of the population from which the lines were formed. Consequently, \( \mathbb{E} \sigma_g^2 = \frac{1}{2} f \sigma_0^2 \) if the foregoing assumptions are correct.

\( m_j \) is an effect, in addition to the \( g_j \) effect, common to all litters having the \( j^{th} \) line as the female parent. As the number of lines become large and all of the possible reciprocal crosses are made in large and equal numbers, the mean of the progeny of the \( j^{th} \) line used as the female parent minus the mean of the progeny of the \( j^{th} \) line used as the male parent approaches \( m_j \). The pre-natal and post-natal mothering ability of a line is measured by \( m_j \). It is a function of the genotype of the line rather than of genes transmitted to the progeny of females of the line. It is assumed that
the $m_j$ are normally distributed with mean $= 0$ and variance $= \sigma_m^2$. Since the lines are assumed to be randomly drawn from the population of lines, $E_{m_j} = 0$. It can be assumed either that the $s_{ij}$ and $m_j$ are uncorrelated or that they are correlated either negatively or positively. If it is assumed that the correlation is not 0, $E_{s_{ij}m_j} = \sigma_{sm}$. Methods were developed in this study for estimating this covariance.

$s_{ij}$ is an effect common to the progeny of matings of the $i^{th}$ line of sire by the $j^{th}$ line of dam and the $j^{th}$ line of sire by the $i^{th}$ line of dam. It is an effect in addition to the $g_i$ and $m_j$ effects. $s_{ij}$ is the measure of how much better or poorer are the progeny of the mean of matings of $i \times j$ and $j \times i$ than would be expected on the basis of exact knowledge of the additive genetic values and maternal values of the lines. Assuming again that the number of lines are large and that all possible reciprocal crosses are made in equal and large numbers, $\frac{1}{2} (\bar{y}_{ij} + \bar{y}_{ji} - \bar{y}_{i} - \bar{y}_{j} - \bar{y}_{.i} - \bar{y}_{.j}) + \bar{y}_{..}$ approaches $s_{ij}$. The dots in the subscripts denote summation and the bars over the y's denote the means. It is assumed that the $s_{ij}$ are normally distributed with mean $= 0$ and variance $= \sigma_s^2$. It must also be assumed that the sum of the $s_{ij}$ over each line $= 0$ because otherwise the $s_{ij}$ and the $g_i$ effects cannot be separated. If all gene action were additive, all of the $s_{ij}$ would $= 0$, and consequently the value of $\sigma_s^2$ would be 0.

Either dominance or epistasis or a combination of the two can cause specific effects, but the methods developed in this study are not adequate to estimate the relative importance of the two. Furthermore, it is not possible to make any statement concerning the expected size of $\sigma_s^2$. This is true because the variances for dominance and epistasis are not known for the original population, and no mathematical statement has been developed for estimating what
effect inbreeding has on the size of the dominance and epistatic variances and on the distribution of these variances among and within inbred lines.

\( r_{ij} \) is an effect common to all progeny of the matings of males of the \( i^{th} \) line to females of the \( j^{th} \) line. It is an effect in addition to the additive genetic, maternal, and specific effects. \( r_{ij} \) is a measure of the difference between reciprocal crosses after account has been taken of the difference in maternal ability between the \( i^{th} \) line and the \( j^{th} \) line. If the number of lines is large, all possible reciprocal crosses are made in large numbers, and the number of experimental units in each of the reciprocal crosses is equal, \( \frac{1}{2} (\bar{y}_{ij} - \bar{y}_{ji} - \bar{y}_{i.} + \bar{y}_{j.} + \bar{y}_{i..} - \bar{y}_{j..}) \) approaches \( r_{ij} \). It is assumed that the \( r_{ij} \) are normally distributed with mean = 0 and variance = \( \sigma_r^2 \). It is also assumed that the sum of the \( r_{ij} \) over each line of dam and over each line of sire is 0. The biological factor which would cause \( r_{ij} \) and \( \sigma_r^2 \) to be real is sex linkage.

In most sets of data there are certain extraneous factors of which account should be taken. Some of these in the case of swine data are inbreeding of the progeny, inbreeding of the dam, age of the dam, and season in which the pig or litter was born. The simplest way to take into account such differences is to adjust the data by means of suitable correction factors. Since, however, few such correction factors are available for swine, it is often necessary to obtain correction factors from the set of data at hand. Additive correction factors can be computed and applied by adding \( a_k \) to the linear mathematical model. \( a_k \) is an effect common to all pigs falling in the \( k^{th} \) class. The \( a_k \) are assumed to be constants, and their sum is assumed to be 0.

\( e_{ijkl} \) is an effect peculiar to the \( ijk^{th} \) experimental unit. It is
assumed to be normally and independently distributed with mean \( \mu \) and variance \( \sigma^2 \). This error variance includes errors of Mendelian sampling, failure of the mathematical model to fit perfectly the actual biology of the material, and a multitude of environmental factors which cannot be measured.

B. Statistical Problems in the Analysis of the Single Cross data

Estimation of the various line and cross line effects, tests of hypotheses, and estimation of the variances present some problems which cannot be solved by the routine techniques of the analysis of variance. This would be true even though the design were a perfectly balanced one in which the same number of experimental units were available for each of the possible reciprocal crosses and the crosses were orthogonal to extraneous factors such as age, inbreeding, and season.

The difficulty is that the diagonal elements with respect to a two-way classification of subclass numbers, line of sire by line of dam are zero. Therefore, the data are not orthogonal. The consequences of non-orthogonality are that the various marginal means give biased estimates of the line and cross line effects, the conventional analysis of variance technique does not give appropriate mean squares for testing hypotheses, and variances cannot be estimated by routine methods. If the design is not a balanced one, the inaccuracies due to applying the usual methods may become very large. The data available for this study were quite unbalanced since the number of litters for the various possible reciprocal crosses varied from 0 to 12 and the crosses were not orthogonal to such factors as age and inbreeding.

The method of least squares provides a means for circumventing the difficulties inherent in non-orthogonality. By properly utilizing the method,
unbiased estimates of the various effects can be obtained, sampling errors of the estimates can be computed, and tests of hypotheses can be affected. Furthermore, estimates of the variances can be obtained from the sums of squares computed by the least squares method. If, however, the different effects are assumed to be randomly drawn from different populations of effects, the least squares estimates of these effects are not the best ones obtainable. Instead the best estimates are the least squares estimates adjusted in a manner described in Section II D. Since in this study the 12 inbred lines are assumed to be randomly drawn from some population of lines, the different effects can also be assumed to be randomly drawn from populations of effects. Consequently, the least squares estimates are not the best estimates. The least squares method was used in this study to test hypotheses, to compute sums of squares from which estimates of variances could be obtained, and to correct various marginal means in accordance with the various criteria of classification accompanying such means. The discussion which follows with respect to the method of least squares, estimation of components of variance, and adjustment of least squares estimates utilizes a more general mathematical model than the model describing the single cross data. This procedure was followed in order to show that the methods are general and because the methods can thereby be more easily described, four rather unusual features being inherent in the single cross model. These unusual features are (1) each experimental unit has in it two different $g_i$, (2) the interaction effect, $a_{ij}$, has interchangeable subscripts, (3) $g_j$ and $m_j$ always go together, and (4) $r_{ij}$ is always accompanied by $a_{ij}$. 
C. The Least Squares Analysis of Non-Orthogonal Data

Brandt (1933) and Yates (1933, 1934) were the first to publish on the least squares method of analysis of multiple classifications with disproportionate subclass numbers. Brandt's method was restricted to a 2 x n classification. Yates extended the analysis to a p x q classification and presented the general theory of tests of hypotheses and the computation of sampling errors. Wilks (1938) and Hazel (1946) described the least squares analysis of a more than two-way classification and introduced an independent variable such as appears in the analysis of covariance. The method of analysis of non-orthogonal data is presented here in detail in order to introduce the notation and to describe the methods developed in this study for adjusting least squares estimates when the effects are assumed to be randomly drawn.

Assume that the mathematical model describing the particular situation is:

\[ y_{ijk} = \mu + a_i + b_j + ab_{ij} + c x_{ijk} + e_{ijk} \]

\[ i = 1, \ldots, p \]
\[ j = 1, \ldots, q \]
\[ k = 1, \ldots, n_{ij} \]

\( x_{ijk} \) is an independent variable associated with \( y_{ijk} \), the dependent variable.

It will first be assumed that \( \mu, a_i, b_j, ab_{ij}, \) and \( c \) are all constants, that \( x_{ijk} \) is measured without error, and that \( e_{ijk} \) is a random variable from a population with mean = 0 and variance = \( \sigma^2 \). The problem is to estimate the various parameters and to test various hypotheses concerning them. In a
model of this sort the least squares method is usually the method of choice because of several desirable properties, namely:

1. The estimates are unbiased, that is, $E(\hat{\theta}) = \theta$, where $\theta$ is the parameter being estimated and $\hat{\theta}$ is the least squares estimate of $\theta$.

2. The sampling error for each estimate is as small or smaller than any other estimate which can be obtained by taking linear combinations of the sample values.

3. The computations can always be carried out.

4. The method provides a straightforward way for obtaining the variance-covariance matrix of the parameter estimates.

5. Parameter estimates can be made independently of any assumption with regard to the distribution of $e_{ijk}$.

6. If the $e_{ijk}$ are assumed to be normally distributed, tests of hypotheses can be effected using the F distribution, and confidence intervals can be computed. Furthermore, the least squares estimates are identical to the maximum likelihood estimates, and the tests of significance are identical to the likelihood ratio tests.

7. The method provides a means for obtaining the maximum amount of information from a set of data with disproportionate subclass frequencies.

1. **Principles of least squares and the normal equations**

   The principle of least squares is that parameter estimates are obtained in such a way as to minimize the sum of the squared residual errors. In the model given above the least squares estimates of $\mu$, $a_i$, $b_j$, $ab_{ij}$, and $c$ are
the set of values which minimize \( \sum_{ijk} e_{ijk}^2 = \sum (y_{ijk} - \mu - a_i - b_j - ab_{ij} - cx_{ijk})^2 \). This is accomplished by differentiating \( \sum e_{ijk}^2 \) partially with respect to each of the parameters, setting each of these derivatives equal to zero, and solving the resulting set of simultaneous equations. The following set of equations is obtained in the present example:

\[
\begin{align*}
\mu: \; n_{..} \mu = & \sum_i n_{i..} a_i + \sum_j n_{i..} b_j + \sum_{ij} n_{ij} ab_{ij} + cx_{..} = y_{..} \\
a_i: \; n_{i..} (\mu + a_i) = & \sum_{ij} n_{ij} (b_j + ab_{ij}) + cx_{i..} = y_{i..} \\
b_j: \; n_{..j} (\mu + b_j) = & \sum_i n_{ij} (a_i + ab_{ij}) + cx_{..j} = y_{..j} \\
ab_{ij}: \; n_{ij} (\mu + a_i + b_j + ab_{ij}) = & cx_{ij} = y_{ij} \\
c: \; x_{..} \mu = & \sum_i x_{i..} a_i + \sum_j x_{..j} b_j + \sum_{ij} x_{ij} ab_{ij} + \sum_{ijk} x_{ijk} = \sum_{ijk} x_{ijk} y_{ijk}
\end{align*}
\]

A parameter followed by a colon denotes equation for that parameter.

2. Estimation of parameters and tests of hypotheses

As the equations stand, no solution is obtainable since the equations are not independent. For example, the sum of the \( a_i \) equations = the sum of the \( ab_{ij} \) equations = the \( \mu \) equation. Similarly, the sum of the \( b_j \) equations = the \( a_i \) equation, and the sum of the \( ab_{il} \) equations = the \( b_1 \) equation. If the following restrictions are now imposed on the estimates, a unique solution to the equations can be obtained; \( \sum a_i = \sum b_j = \sum ab_{ij} \) (for all \( j \)) = \( \sum ab_{ij} \) (for all \( i \)) = 0. These restrictions reduce the number of unknowns from \( p + q + pq + 2 - m \) to \( pq + 1 - m \), where \( m \) is the number of subclasses in which \( n_{ij} = 0 \). The number of degrees of freedom is also \( pq + 1 - m \).
including p-1 for A, q-1 for B, (p-1)(q-1) - m for A x B, one for μ, and one for c. The first hypothesis to be tested is that \( ab_{ij} = 0 \), for on the outcome of that test depends further steps in the procedure. To test this hypothesis the sums of squares for A x B and for error are needed. They are computed as follows:

\[
(A \times B)\text{S.S.} = R(\mu, a, b, ab, c) - R(\mu, a, b, c).
\]

\[
\text{Error S.S.} = R(T) - R(\mu, a, b, ab, c).
\]

\( R(\ ) \) denotes the reduction in sum of squares due to fitting the particular parameters shown inside the parentheses.

\( R(T) \) denotes \( \sum y_{ijk}^2 \), that is, the total sum of squares without the correction factor removed. The reduction due to fitting all of the parameters is \( R(\mu, a, b, ab, c) = \hat{\mu} y_{..} + \sum \hat{a}_i y_{i..} + \sum \hat{b}_j y_{.j} + \sum \hat{ab}_{ij} y_{ij} + \hat{c} \sum x_{ijk} y_{ijk} \).

The \( \hat{\ } \) over a parameter denotes the least squares estimate under the unrestricted hypothesis.

\[
R(\mu, a, b, c) = \hat{\mu} y_{..} + \sum \hat{a}_i y_{i..} + \sum \hat{b}_j y_{.j} + \hat{c} \sum x_{ijk} y_{ijk}.
\]

The \( \sim \) denotes new least squares estimates made under the hypothesis that the \( ab_{ij} = 0 \). The latter estimates are made from the original set of equations from which the \( ab_{ij} \) equations and unknowns are deleted.

The computation of R(\( \mu \), a, b, ab, c) is most easily accomplished by letting \( \tilde{ab}_{ij} = \mu + a_i + b_j + ab_{ij} \). Then \( R(\mu, a, b, ab, c) = \sum \tilde{ab}_{ij} y_{ij} + \hat{c} \sum x_{ijk} y_{ijk} \).

\( \tilde{ab}_{ij} \) is easily estimated if use is made of only the \( ab_{ij} \) and c equations. There are \( pq + 1 - m \) such equations, and, as stated above, that is the number of degrees of freedom for these parameters. The equations to be solved are, therefore,
the familiar within subclass regression of analysis of covariance. Substituting \( \hat{c} \) for \( c \) in the \( \tilde{a}_{ij} \) equations, 

\[
\hat{a}_{ij} = \frac{1}{n_{ij}} \left( y_{ij} - x_{ij} \hat{c} \right).
\]

In the computation of \( R(\mu, a, b, c) \), \( \mu \) can be combined with either \( a_i \) or \( b_j \) whichever is the more numerous to give say \( a_i = \mu + a_i \). The unreduced equations then become

\[
\tilde{a}_i = n_{i..} \tilde{a}_i + \sum_j n_{ij} a_j + x_{i..} c = y_{i..}
\]

\[
b_j = \sum_i n_{ij} \tilde{a}_i + n_{..j} b_j + x_{..j} c = y_{..j}
\]

\[
c = \sum_i x_{i..} \tilde{a}_i + \sum_j x_{..j} b_j + c \sum x_{ijk} = \sum x_{ijk} y_{ijk}
\]

Now since 

\[
\tilde{a}_i = \frac{1}{n_{i..}} \left( y_{i..} - \sum_j n_{ij} b_j - x_{i..} c \right),
\]

and the \( b_j \) equations can be reduced to ones involving only \( b_j \) and \( c \), the solution of which after introducing \( \sum b_j = 0 \), yields \( \hat{b}_j \) and \( \hat{c} \). Then

\[
\hat{a}_i = \frac{1}{n_{i..}} \left( y_{i..} - \sum_j n_{ij} \hat{b}_j - x_{i..} \hat{c} \right).
\]

The computational procedure for absorbing the coefficients of one set of constants is given in more detail in Section II C 3 a. The \( A \times B \) sum of squares calculated in the manner just described has \( (p-1)(q-1) - m \) degrees of freedom, and the error sum of squares has \( n_{...} - pq + m \) degrees of freedom. The ratio of the two mean squares is distributed as \( F \) if the e_{ijk} are
normally distributed. The foregoing procedure illustrates the general method of computing sums of squares. That is, the sum of squares for a particular set of parameters is the reduction due to fitting all of the parameters minus the reduction due to fitting all of the parameters except those in question.

Ordinarily the real purpose of an analysis is the estimation of the parameters. In the example just given no estimates of $a_1$, $b_j$, $ab_{ij}$, and $\mu$ as such were obtained in the computation of $R(\mu, a, b, ab, c)$. These estimates can, however, be obtained from the $\hat{ab}_{ij} = \hat{\mu} + \hat{a}_1 + \hat{b}_j + \hat{ab}_{ij}$. Assuming that all of the AB subclasses are filled, the estimates are as follows,

$$\hat{\mu} = \frac{1}{pq} \sum_{ij} \hat{ab}_{ij}$$

$$\hat{a}_1 = \frac{1}{q} \sum_{j} \hat{ab}_{ij} - \hat{\mu}$$

$$\hat{b}_j = \frac{1}{p} \sum_{i} \hat{ab}_{ij} - \hat{\mu}$$

$$\hat{ab}_{ij} = \hat{ab}_{ij} - \hat{\mu} - \hat{a}_1 - \hat{b}_j$$

If some of the subclasses are missing, it is necessary to solve the following equations in order to estimate $\mu$, $a_1$, and $b_j$:

$$\mu: n \cdot \mu + \sum_i n_{i1}a_i + \sum_j n_{j}b_j = \hat{ab}_{..}$$

$$a_1: n_{i1} \mu + n_{i1}a_i + \sum_j n_{ij}b_j = \hat{ab}_{i1}$$

$$b_j: n_{j} \mu + \sum_i n_{ij}a_i + n_{j}b_j = \hat{ab}_{.j}$$

$n_{ij} = 1$ or 0 depending upon whether the $ij^{th}$ subclass is filled. If $ab_{ij} \neq 0$, the tests of hypotheses that $a_1 = 0$ and that $b_j = 0$ are
very tedious to compute unless all subclass are filled. Formally, the
S.S. for \( A = R(\mu, a, b, ab, c) - R(\mu, b, ab, c) \). The difficulty is that
there appears to be no way for absorbing the coefficients of the \( ab_{ij} \). A
special method for setting up the equations is also required.

If the hypothesis that \( ab_{ij} = 0 \) is accepted, the sum of squares for \( A
= R(\mu, a, b, c) - R(\mu, b, c) \), and that for \( B = R(\mu, a, b, c) - R(\mu, a, c) \).
The computation of \( R(\mu, a, b, c) \) was described above. Letting \( \hat{b}_j = \mu + b_2 \),
the equations for \( R(\mu, b, c) \) are,

\[
\sum b_j: n \cdot \hat{b}_j + x \cdot j \cdot c = y \cdot j.
\]

\[
\sum x \cdot j \cdot \hat{b}_j + c \sum x \cdot j \cdot 2 = \sum x \cdot j \cdot y \cdot j \cdot \hat{y}_j.
\]

It turns out that \( \hat{c} = \sum x \cdot j \cdot y \cdot j \cdot \hat{y}_j - \frac{\sum x \cdot j \cdot y \cdot j\cdot \hat{y}_j}{n \cdot j} \), which is the within
B regression of \( y \) on \( x \). Then \( \hat{b}_j = \frac{1}{n \cdot j} (y \cdot j - x \cdot j \cdot \hat{c}) \), and \( R(\mu, b, c) = \)
\[
\frac{\sum \hat{b}_j \cdot y \cdot j\cdot \hat{c} \sum x \cdot j \cdot y \cdot j \cdot \hat{y}_j. \]

The error sum of squares is \( R(T) - R(\mu, a, b, c) \).

If it is desired to test the hypothesis that \( c = 0 \) under the assump-
tion that \( a_4, b_4, ab_{ij} \neq 0 \), the appropriate sum of squares is \( R(\mu, a, b, ab,
c) - R(\mu, a, b, ab) \). The first of these reductions has already been de-
scribed. The equations for the second can be written,

\[
\sum b_{ij}: n_{ij} \hat{b}_{ij} = y_{ij}.
\]

Therefore, \( \hat{c} \sum \hat{b}_{ij} = \frac{1}{n_{ij}} y_{ij} \), and \( R(\mu, a, b, ab) = \frac{1}{n_{ij}} y_{ij}^2 \), which is
just the usual subclass sum of squares before the correction factor is sub-
tracted. The error sum of squares is \( R(T) - R(\mu, a, b, ab, c) \).

The hypothesis that \( c = 0 \) can also be tested assuming \( ab_{ij} = 0 \). The
sum of squares for this case is \( R(\mu, a, b, c) - R(\mu, a, b) \).

Different estimates of a given set of parameters are obtained, depending upon what assumptions are made concerning the remaining parameters in the mathematical model. Similarly, different sums of squares are obtained depending upon the assumptions made. It is therefore necessary to follow some logical procedure with respect to order of tests. For example, first test for significance the estimate of the parameter or set of parameters which seems least likely to be significant. If the hypothesis that the set = 0 is accepted, proceed to test the set thought next least likely to be significant under the assumption that the first set = 0. If the first hypothesis had been rejected, the set of parameters involved in that test would have been retained in testing the second set. In a like manner, proceed to test each set of estimates, one at a time, in reverse order of their probable significance, deleting from subsequent tests each set which does not appear to be real. The reason for deleting sets of parameters not found to be of consequence is that the tests become less powerful and the sampling errors of estimates become larger as more parameters are estimated in a particular set of non-orthogonal data. Unfortunately, there is no "rule of thumb" method for deciding in what order to make the tests or at what level of significance to reject an hypothesis and consequently to retain a set of parameters in subsequent tests. This problem does not arise in orthogonal data since the least squares estimates for each set of parameters are obtained independently of the other parameters.

3. Computational shortcuts in the solution of least squares equations

The difficulty in solving large sets of simultaneous equations has been
a serious deterrent to the use of least squares analyses of non-orthogonal data. Utilization of two computational procedures, the absorption of the coefficients of one set of the parameters into the others and the iterative solution of the reduced equations, greatly reduce the time required in carrying out a least squares analysis. These two procedures are described below.

a. Absorption of the coefficients of one set of parameters. Much labor in solving simultaneous equations and in inverting least squares matrices can be saved by utilizing the fact that \( \mu \) and the parameters associated with one classification can by some simple algebra be absorbed into the coefficients of the other parameters. This procedure was described for the two way classification by Yates (1934). If there are \( p \) A classes, \( q \) B classes, and \( r \) C classes and if \( \mu \) and the \( a_i \) coefficients are absorbed, the number of equations is reduced from \( p + q + r - 2 \) to \( q + r - 2 \). The method is based on the fact that \( \mu \) can be combined with \( a_i \) to form the parameters \( \mu + a_i \). The method will be illustrated with the mathematical model,

\[
y_{ijkl} = \mu + a_i + b_j + c_k + d x_{ijkl} + e_{ijkl}
\]

The coefficients of the least squares equations are given below.

\( \bigcirc \) denotes diagonal elements.

The \( \bigcirc \) denotes that all other entries off the diagonal in that particular block are 0. The right members of the equations are given in the column with heading "Sum".

\[
i = 1, \ldots, p
\]
\[
j = 1, \ldots, q
\]
\[
k = 1, \ldots, r
\]
\[
l = 1, \ldots, n_{ijk}
\]
Since $\mu + a_i = \frac{1}{n_{i..}} (y_{i..} - \sum_j n_{ij} b_j - \sum_k n_{ik} c_k - d x_{i..})$, the equations can be written in a form containing only $b_j$, $c_k$, and $d$, thus:

<table>
<thead>
<tr>
<th>$b_j$</th>
<th>$c_k$</th>
<th>$d$</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_j$ : $C(b, b_j)$</td>
<td>$C(b_j, c_k)$</td>
<td>$C(b_j, d)$</td>
<td>$S(b_j)$</td>
</tr>
<tr>
<td>$c_k$ : $C(b, b_j)$</td>
<td>$C(c_k, c_k)$</td>
<td>$C(c_k, d)$</td>
<td>$S(c_k)$</td>
</tr>
<tr>
<td>$d$ : $C(b_j, d)$</td>
<td>$C(c_k, d)$</td>
<td>$C(d)$</td>
<td>$S(d)$</td>
</tr>
</tbody>
</table>

The $C(\quad )$'s denote new coefficients of the equations, and the $S(\quad )$'s denote new right members of the equations. The new coefficients and sums are obtained as follows:

$$C(b_j b_j) = n_{..j} - \sum_i n_{ij} \cdot \frac{2}{n_{..i}}$$

$$C(b_j b_j') = -\sum_{j \neq j'} n_{ij} \cdot \frac{n_{ij'}}{n_{..i}}$$

$$C(b_j c_k) = n_{..jk} - \sum_i n_{ij} \cdot \frac{n_{ik}}{n_{..i}}$$
\[ C(b_j d) = x_{j..} - \sum_i n_{ij} x_{i...}/n_{i...} \]
\[ S(b_j) = y_{j..} - \sum_i n_{ij} y_{i...}/n_{i...} \]
\[ C(c_k c_k) = n_{..k} - \sum_i n_{i..}^2/k/n_{i...} \]
\[ C(c_k c_{k'}) = \sum_{i,i'} n_{i..} n_{i'..}/n_{i...}' \]
\[ C(c_k d) = x_{..k..} - \sum_i n_{i..} x_{i...}/n_{i...} \]
\[ S(c_k) = y_{..k..} - \sum_i n_{i..} y_{i...}/n_{i...} \]
\[ C(d) = x_{ijkl}^2 - \sum_i x_{i...}^2/n_{i...} \]
\[ S(d) = x_{ijkl} y_{ijkl} - \sum_i x_{i...} y_{i...}/n_{i...} \]

A useful check on the computations is provided by the fact that \( \sum_{j=1}^j \)
\[ C(b_a b_j) = \sum_{k=1}^k C(b_j c_k) = \sum_{k=1}^k C(c_k c_k) = \sum_{j=1}^j C(b_j d) = \sum_{k=1}^k C(c_k d) = \sum_{j=1}^j S(b_j) \]
\[ = \sum_{k=1}^k S(c_k) = 0. \]

Since each of the \( n_{i...} \) appears as a divisor in every expression for
the new coefficients, the computations can be carried out most conveniently
by setting up columns for all the \( b_j, c_k, \) and \( d \) as follows:

<table>
<thead>
<tr>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>...</th>
<th>( c_1 )</th>
<th>( c_2 )</th>
<th>...</th>
<th>( d )</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>( n_{11}/n_{1..} )</td>
<td>( n_{12}/n_{1..} )</td>
<td>...</td>
<td>( n_{1.1}/n_{1..} )</td>
<td>( n_{1.2}/n_{1..} )</td>
<td>...</td>
<td>( x_{1...}/n_{1...} )</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>( n_{21}/n_{2..} )</td>
<td>( n_{22}/n_{2..} )</td>
<td>...</td>
<td>( n_{2.1}/n_{2..} )</td>
<td>( n_{2.2}/n_{2..} )</td>
<td>...</td>
<td>( x_{2...}/n_{2...} )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Then the elements of these columns can be multiplied by the corresponding elements of the original equations and the sums of these products accumulated in the calculating machine. After these computations have been effected, \( q + r + 1 \) equations remain, but only \( q + r - 1 \) of the parameters can be estimated independently. Consequently, use is made of the fact that \( \sum b_j = \sum c_k = 0 \) to subtract in each equation the coefficient of \( b_q \) from the coefficients of each of the other \( b_j \) and the coefficient of \( c_r \) from the coefficients of each of the other \( c_k \). The \( b_q \) and \( c_r \) equations are then deleted. The solution to these equations gives the least squares estimates of the first \( q - 1 \) of the \( b_j \), of the first \( r - 1 \) of the \( c_k \), and of \( \delta \). Then \( b_q = - \frac{1}{q-1} \sum_{j=1}^{q-1} b_j \) and \( c_r = - \frac{1}{r-1} \sum_{k=1}^{r-1} c_k \). Substitution of these estimates in the \( \mu + a_1 \) equations of the original set gives the least squares estimate of \( \mu + a_1 \). If \( \hat{\mu} \) and the \( \hat{a}_1 \) are wanted separately, \( \hat{\mu} = \frac{1}{p} \sum \hat{\mu} + \hat{a}_1 \) and \( \hat{a}_1 = (\hat{\mu} + \hat{a}_1) - \hat{\mu} \).

b. **Iterative solution of the equations.** Although absorption of the coefficients of one set of the parameters into the coefficients of the other parameters materially reduces the number of equations, the number requiring simultaneous solution may still be very large. If the number exceeds 10, a solution by the usual procedure of eliminating one unknown at a time becomes very laborious, even when advantage is taken of the possibility for maintaining the symmetry of the equations step by step. Under most circumstances, the least squares equations can be solved much more easily by an iterative process such as the following described one.

Assume the following set of equations:
The \( a_j \) are parameters to be estimated, the \( a_1 \) are equations obtained by differentiating \( \sum e^2 \) with respect to \( a_1 \), and the \( y_i \) are the right members of the equations. The iterative method involves the obtaining of successive estimates of the \( a_j \), say \( \hat{a}_j, \tilde{a}_j, \ldots, \tilde{a}_j, \ldots \). The first step in the procedure is the making of first guesses as to the values of the \( \tilde{a}_j \). The marginal means frequently serve as a good starting point. If these first guesses are called \( \hat{a}_j \), then

\[
\hat{a}_1 = \frac{1}{C_{11}} \left[ y_1 - \sum_{j \neq 1} C_{1j} \hat{a}_j \right]
\]

\[
\hat{a}_2 = \frac{1}{C_{22}} \left[ y_2 - C_{21} \hat{a}_1 - \sum_{j \neq 1,2} C_{2j} \hat{a}_j \right]
\]

\[
\hat{a}_s = \frac{1}{C_{ss}} \left[ y_s - \sum_{j=1}^{s-1} C_{sj} \hat{a}_j - \sum_{j=s+1}^{p} C_{sj} \tilde{a}_j \right],
\]

where \( s \) denotes any particular \( j \leq p \). If this process is continued until all \( \hat{a}_j = \tilde{a}_j \), the \( \tilde{a}_j \) are the solution to the set of simultaneous equations.

c. Solutions utilizing the inverse matrix. If the inverse of the

\[
\begin{array}{cccccc}
  a_1 & a_2 & \cdots & a_j & \cdots & a_p \\
  \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
  a_1 & C_{11} & C_{12} & \cdots & C_{1j} & \cdots & C_{1p} & y_1 \\
  a_2 & C_{21} & C_{22} & \cdots & C_{2j} & \cdots & C_{2p} & y_2 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  a_1 & C_{11} & C_{12} & \cdots & C_{1j} & \cdots & C_{1p} & y_1 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  a_p & C_{pl} & C_{p2} & \cdots & C_{pj} & \cdots & C_{pp} & y_p \\
\end{array}
\]
matrix of coefficients of the least squares equations is available, it can be used in obtaining the solution of the least squares equations. This procedure is particularly advantageous in cases where different sets of observations are taken on the same experimental material. For example, in this study the litter was used as the experimental unit, and 8 different measurements were studied, number of pigs at birth, 21 days, 56 days, and 154 days and litter weights at the same ages. Eight separate sets of solutions of the simultaneous equations are required to obtain the parameter estimates, but one solution of the inverse matrix can be used for all eight.

Say that the model is \( y_{ijk} = \mu + a_i + b_j + e_{ijk} \) and the elements of the inverse matrix are \( \mu^{-1}, \mu^{-1}a_i, \mu^{-1}b_j, a_i a_i, a_i b_j, b_j b_j, \) and \( b_j b_j' \). Then,

\[
\hat{\mu} = \sum_i \mu^{-1}a_i y_{i..} + \sum_j \mu^{-1}b_j y_{.j}.
\]

\[
\hat{a}_i = \sum_i a_i y_{i..} + \sum_j a_i b_j y_{.j}.
\]

\[
\hat{b}_j = \sum_i b_j y_{i..} + \sum_j b_j b_j y_{.j}.
\]

In case the inverse matrix involves \( \hat{a}_i = \mu + a_i \), the estimates are

\[
\hat{a}_i = \sum_i a_i y_{i..} + \sum_j a_i b_j y_{.j}.
\]

\[
\hat{b}_j = \sum_i b_j y_{i..} + \sum_j b_j b_j y_{.j}.
\]

The inverse matrix and its computation are described in Section II C 4.

4. **Sampling errors of least squares estimates**

   a. **The variance-covariance matrix.** The variance-covariance matrix, \( \sigma a_i a_j \), of least squares estimates is equal to \( a_i a_j / \sigma_2^2 \), where \( a_i a_j \) is the inverse of the matrix of coefficients of the least squares
equations, and $\sigma_e^2$ is the mean square for error. The matrix of coefficients of the usual least squares equations has a zero determinant and consequently cannot be inverted. This comes about because the sum of the equations pertaining to each set of parameters is identical to the $\mu$ equation. It is, therefore, necessary to reduce the equations to the number of independent parameters (degrees of freedom). The mathematical model

$$y_{ijkl} = \mu + a_i + b_j + c_k + d x_{ijkl} + e_{ijkl}$$

is again used as an illustration. The least squares equations were given in Section II C 3 a. This matrix is $(p + q + r + 1)^2$ in size and needs to be reduced to $(p + q + r - 1)^2$, the degrees of freedom, which include 1 for $\mu$, $p - 1$ for $a_i$, $q - 1$ for $b_j$, $r - 1$ for $c_k$, and 1 for $d$.

Therefore, if use is made of the fact that $\sum \hat{b}_j = \sum \hat{c}_k = 0$ to eliminate $b_j$ and $c_k$, the unknowns can be reduced to the appropriate number. This procedure involves subtracting the coefficient of $b_q$ from the other $b_j$ coefficients and the coefficient of $c_r$ from the coefficients of the other $c_k$ in all of the $p + q + r + 1$ equations. Then, in order to reduce the number of equations to $p + q + r - 1$ while maintaining the symmetry of the matrix of coefficients, the $b_q$ equation is subtracted from each of the other $b_j$ equations, and the $c_r$ equation is subtracted from each of the other $c_k$ equations. The reduced matrix is shown below.

This matrix can now be inverted. Let the elements of the invert matrix be denoted by the row and column parameters with accompanying superscripts rather than subscripts. For example, $d_{jck}$ denotes the element of the $b_j$ row and $c_k$ column (or the $c_k$ row and $b_j$ column). An invert element with respect to $d$ is denoted by $d^{-1}$ and with respect to $\mu$ by $\mu^{-1}$. Since the least squares matrix is symmetric about the main diagonal, the invert matrix
is similarly symmetric. The diagonal elements multiplied by \( \sigma_e^2 \) are the variances, and those off the diagonal multiplied by \( \sigma_e^2 \) the covariances.

\[
\begin{array}{cccc}
\mu + a_i & b_j & c_k & d \\
\hline
\mu + a_i & n_{ij} - n_{iq} & n_{i,k} - n_{i,r} & x_{i...} \\
\hline
b_j & n_{ij} - n_{iq} & n_{jr} + n_{qr} & x_{j...} - x_{q...} \\
\hline
c_k & n_{i,k} - n_{i,r} & n_{jr} + n_{qr} & x_{k...} - x_{r...} \\
\hline
d & x_{i...} & x_{j...} - x_{q...} & x_{k...} - x_{r...} + \sum x_{ijkl}^2 \\
\end{array}
\]

The elements of the inverse matrix pertaining to \( b_q \) and \( c_r \) are easily obtained by utilizing these relationships,

\[
\begin{align*}
 a_{1q} &= - \sum_{j=1}^{k} a_{1j} b_{jq}, & a_{1r} &= - \sum_{k=1}^{l} a_{1k} c_{rk}, \\
 b_{sj} &= - \sum_{j=1}^{k} b_{sj} b_{jq}, & b_{sq} &= - \sum_{j=1}^{k} b_{jq} b_{sq}, & b_{sk} &= - \sum_{j=1}^{k} b_{jq} b_{sk}, \\
 b_{sd} &= - \sum_{j=1}^{k} b_{jd} d_{sj}, & c_{rj} &= - \sum_{k=1}^{l} c_{rk} c_{rk}, & c_{rk} &= - \sum_{k=1}^{l} c_{rk} c_{rk}, \\
 c_{rd} &= - \sum_{k=1}^{l} d_{kd} d_{rk}. &
\end{align*}
\]

In the discussions which follow with regard to the invert matrix, it will be understood that the matrix includes the elements pertaining to the dependent parameters.

b. Inversion of the matrix of coefficients. A number of ways have been suggested for inverting matrices like those arising in least squares
analyses. Most of the methods take advantage of the symmetry of the equations. Dwyer has recently discussed these methods (1941). It was found in this study, however, that the inversions could be most quickly accomplished by an iterative method. Two important time-saving adjuncts to the method will also be described. These are (1) the absorption of one set of coefficients and (2) the setting up of a non-symmetric matrix for computational purposes.

(1) Absorption of the coefficients of one set of parameters.

If the coefficients of one set of parameters are absorbed into the coefficients of the remainder as previously described for solution of simultaneous equations arising in least squares analyses (Section II C 3 a), the elements of the inverse of the reduced matrix are identical with the corresponding block of the inverse matrix of the entire set of coefficients in the original least squares equations. It is then a matter of simple algebra to calculate the remaining elements of the inverse matrix. They are:

\[
\bar{a}^s_{m} = -\frac{1}{n_{s..}} (\sum_{j} n_{s.j} b^t b^j + \sum_{k} n_{s.k} b^t c^k + x_{s} \cdots b^d^{-1})
\]

\[
\bar{a}^s_{c} = -\frac{1}{n_{s..}} (\sum_{j} n_{s.j} b^j c^u + \sum_{k} n_{s.k} c^u c^k + x_{u} \cdots c^u d^{-1}),
\]

and similarly for all other \(a^4 b^j\) and \(a^4 c^k\).

\[
\bar{a}^d = -\frac{1}{n_{s..}} (\sum_{j} n_{s.j} b^d c^j + \sum_{k} n_{s.k} c^d c^k + x_{s} \cdots d^{-1})
\]

and similarly for all other \(a^4 d^{-1}\).

\[
\bar{a}^s a^s = \frac{1}{n_{s..}} (1 - \sum_{j} n_{s.j} \bar{a}^s b^j - \sum_{k} n_{s.k} \bar{a}^s c^k + x_{s} \cdots \bar{a}^s d^{-1}),
\]

and similarly for all other \(a^4 a^4\).

\[
\bar{a}^s a^t = -\frac{1}{n_{t..}} \left( \sum_{i} n_{t.i} a^i b^i + \sum_{i} n_{t.k} a^i c^k - x_{t} \cdots \bar{a}^s d^{-1} \right)
\]
If the sampling errors of \( \hat{\mu} \) and \( \hat{\alpha}_i \) are desired separately, the following equations yield the necessary inverse elements:

\[
\mu^{-1} - 1 = \frac{1}{p^2} \text{ (Sum of all elements in the } a_i \text{ block of the inverse matrix)}.
\]

\[
\mu^{-1} a_s = \frac{1}{p} \sum_i a_s a_i - \mu^{-1} a_s.
\]

\[
a^t_s a = \mu^{-1} a_t - \mu^{-1} a_s - \mu^{-1} a_t - \mu^{-1} a_s.
\]

\[
a^t_s = -\sum_i a_i a_s.
\]

Similarly, \( \mu^{-1} b^j = \frac{1}{p} \sum_i a^j b^i \) and \( a^j = a^j b - \mu^{-1} b^j \).

\[
\mu^{-1} c = \frac{1}{p} \sum_i a^k c \quad \text{ and } \quad a^k = a^k c - \mu^{-1} c.
\]

(2) The iterative method. Assume the same mathematical model as before, that \( \tilde{a}_i = \mu + a_i \) have been absorbed, that \( b_q \) and \( c_r \) have been eliminated by subtracting their coefficients from the \( b_q \) and \( c_r \) respectively, and that the symmetry has been maintained by subtracting the \( b_q \) and \( c_r \) equations from the other \( b_q \) and \( c_r \) equations. Since symmetry is present, the definition of an inverse matrix requires that the sum of the products of the elements of the \( i^{\text{th}} \) row of the matrix by the corresponding elements of the \( i^{\text{th}} \) row of its inverse must equal one. The corresponding sum of the \( i^{\text{th}} \) row by any other row must equal zero. Therefore, if the matrix is equated to \( \begin{pmatrix} 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix} \), the solution to the set of simultaneous equations gives the first row (and the first column) of the invert matrix. Similarly, equating to \( \begin{pmatrix} \mu & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & \mu & 0 \end{pmatrix} \) and solving
gives the second row (column) of the invert matrix. A total of $q + r - 1$ such solutions, the last one having as its right member $\begin{pmatrix} \vdots \\ 0 \end{pmatrix}$, yields the inverse matrix of the $b_j$, $c_k$, and $d$ block. Since the inverse is symmetric, each successive step in the solution permits dropping out one additional equation. For example, after the $b_1$ row has been determined, the $b_1$ equation is deleted since the $b_1$ unknown of the $b_2$ equation is already known. This unknown is $b_1^1 b_2$ obtained in the solution to the first set of equations. After the $b_2$ row has been determined, the $b_1$ and $b_2$ equations are both deleted since the $b_1$ unknown of the $b_3$ equation is $b_1^1 b_3$ obtained in the first solution, and the $b_2$ unknown of the $b_3$ equation is $b_2 b_3$ obtained in the second solution. Each of these sets of equations can be solved by the iterative method previously described. If, however, the iterative method is to be used, a non-symmetric matrix which is much easier to solve iteratively can be utilized.

(3) Use of a non-symmetric matrix. It is desirable in the iterative solution that the diagonal elements of the matrix of coefficients be as large as possible compared to the elements off the diagonal. The method described for retaining symmetry while reducing the equations to the number of independent parameters makes the non-diagonal elements quite large. The method described below avoids this difficulty. Assume that $\mu + a_1$ have been absorbed, and that $b_1$, $b_2$, $c_1$, $c_2$, $d_1$, $d_2$, $\ldots$, $\ldots$,
c_r and d are left. Now delete the b_q and c_r equations, and in each of the remaining equations subtract the coefficient of b_q from the coefficients of the other b_j and the coefficient of c_r from the coefficients of the other c_k. Then equate to the following matrix:

\[
\begin{pmatrix}
 b_1 & b_2 & \ldots & b_{q-1} & c_1 & c_2 & \ldots & c_{r-1} & d \\
 b_1 & (q-1)/q & -1/q & \ldots & -1/q & 0 & 0 & \ldots & 0 & 0 \\
 b_2 & -1/q & (q-1)/q & \ldots & -1/q & 0 & 0 & \ldots & 0 & 0 \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
 b_{q-1} & -1/q & -1/q & \ldots & (q-1)/q & 0 & 0 & \ldots & 0 & 0 \\
c_1 & 0 & 0 & \ldots & 0 & (r-1)/r & -1/r & \ldots & -1/r & 0 \\
c_2 & 0 & 0 & \ldots & 0 & -1/r & (r-1)/r & \ldots & -1/r & 0 \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
c_{r-1} & 0 & 0 & \ldots & 0 & -1/r & -1/r & \ldots & (r-1)/r & 0 \\
d & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 \\
\end{pmatrix}
\]

D. Adjustment of Least Squares Estimates When Effects are Randomly Drawn from Populations With Known Variances

The method of estimation by least squares requires no assumptions concerning the nature of the parameters being estimated or of the nature of the errors. All that is required is that there be a mathematical model and that the parameters be estimated in such a way that the estimates will
minimize the sum of squares for error. If tests of hypotheses are to be
affected and confidence intervals constructed, it is necessary only to as-
sume that the errors are normally distributed. But if certain assumptions
are made concerning the parameters, for example, the $a_i$ are randomly drawn
from a normal population with mean = 0 and variance = $\sigma^2$, the least squares
method of estimating the $a_i$ makes no use whatsoever of this information.
Failure of the least squares method to utilize such information must surely
result in less than maximum efficiency of estimation.

Very many practical statistical problems do involve estimation of cer-
tain parameters concerning which reasonable assumptions can be made. For
example, most of the problems in animal breeding involve estimation of breed-
ing values of individuals, families, or lines. These individuals, families,
or lines are usually assumed to be members, often randomly drawn ones, from
some population concerning which there is a certain amount of knowledge. In
this study estimates of the $g_i$, $m_j$, $a_ij$, and $r_{ij}$ are desired. Since it is
assumed that each of these effects is randomly drawn from a normal popula-
tion, the least squares estimates of the effects fail to make full use of the
assumptions of the mathematical model. Consequently, the least squares esti-
mates are not the best possible ones for this problem. It is convenient to
think of the least squares method as being only a powerful computational pro-
ceedure for obtaining the best possible additive correction factors from a
given set of data classified into two or more criteria of classification and
for automatically applying these correction factors to the means of the vari-
ob classes.

Given the least squares estimate of a parameter, what is the best esti-
mate of the parameter when it is assumed that the parameter is randomly drawn
from a normal population? It is a plausible hypothesis that estimates of above average size contain interaction and error terms whose sum is positive and that estimates of below average size have in them interaction terms and error terms whose sum is negative. It therefore seems logical to regress the least squares estimate toward zero if we wish to obtain the best estimate of the parameter. What is needed to accomplish this is the regression of the parameter on the least squares estimate of the parameter, say \( b(\theta) \). In the usual regression theory \( b(\theta) = \frac{E \theta}{E \theta^2} \). This ratio can always be computed for any particular \( \hat{\theta} \). Then \( \hat{\theta} = b(\theta) \theta \) where \( \hat{\theta} \) denotes best estimate and \( \theta \) denotes the least squares estimate of \( \theta \).

1. The maximum likelihood solution

The important question is whether \( \hat{\theta} \) is really the best estimate of \( \theta \). The maximum likelihood estimate is ordinarily the estimate of choice provided it can be obtained. It will now be shown that under certain reasonable assumptions the maximum likelihood estimate of \( \theta \), given that the least squares estimate of \( \theta \) is \( \hat{\theta} \), is actually \( \hat{\theta} \).

A most general linear model can be used for this proof. Let \( y = \mu + a_i + \) other parameters + error. The \( a_i \) are assumed to be randomly drawn from a normal population with mean = 0 and variance = \( \sigma^2 \). Also, the errors and interactions containing \( A \) are assumed to be normally distributed. No assumptions need be made regarding the other parameters. Given that the least squares estimate, \( \hat{a}_i \), is available, what is the maximum likelihood estimate of \( a_i \)? The joint distribution of \( a_i \) and \( \hat{a}_i \) is \( f(a_i, \hat{a}_i) = g(a_i) h(\hat{a}_i|a_i) \), where \( g(a_i) \) is the marginal distribution of \( a_i \) and \( h(\hat{a}_i|a_i) \) is the conditional distribution of \( \hat{a}_i \) given \( a_i \).
\[ g(a_1) = N(C, Ea_1^2) \]
\[ h(a, \hat{a}_1) = N \left[ a_1, E(\hat{a}_1 | a_1) \right]^2 \]
Then \( \frac{\partial \log L}{\partial a_1} = \frac{-a_1}{Ea_1^2} + \frac{\hat{a}_1 - a_1}{E(\hat{a}_1 | a_1)^2} = 0 \), and
\[ \hat{a}_1 = \frac{Ea_1^2}{Ea_1^2 + E(\hat{a}_1 | a_1)^2}, \]

It is now a question of determining whether the coefficient of \( \hat{a}_1 \) in the maximum likelihood estimate of \( a_1 \) is equal to the regression of \( \hat{a}_1 \) on \( \hat{a}_1 \). If it is assumed that \( \mu \) is known, the expectation of \( \hat{a}_1 \) is \( a_1 + \) interaction terms + error terms. Consequently, \( Ea_1 \hat{a}_1 = Ea_1^2 \) proving that the numerators in the two coefficients are equal. \( E(\hat{a}_1 | a_1)^2 = Ea_1^2 \). Therefore, \( Ea_1^2 + E(\hat{a}_1 | a_1)^2 = Ea_1^2 \) proving that the two denominators are equal. If \( \mu \) is not known, the expectation of \( \hat{a}_1 \), is \( \frac{Ea_1^2}{Ea_1^2 + \frac{1}{p} \sum a_1} \) + interaction terms + error terms. Consequently, the maximum likelihood estimate and the estimate obtained by regressing the least squares estimate differ by a factor of \( \frac{Ea_1^2}{p} \) as compared to 1 with respect to the coefficient of \( Ea_1^2 \) in both the numerator and denominator.

Another method of estimation involves choosing a regression coefficient to apply to the least squares estimate such that the expectation of \( (b \hat{a}_1 - a_1)^2 \) is a minimum. As would be expected, the value of \( b \) which minimizes this expression is \( \frac{Ea_1 \hat{a}_1}{Ea_1^2} \).

In order to show that many of the presently used techniques in animal breeding are special applications of the general theory of estimation just described, two commonly used procedures will be described, namely, the estimation of breeding value on the basis of a single record and estimation of most probable producing ability from \( n \) records. In the first case, if it is assumed that the several individuals made their records in the same herd and under the same general environmental conditions, the most probable
breeding value of an individual is estimated to be the herd average + heritability times the amount by which the record of the individual exceeds the herd average. A mathematical model describing this situation is 

\[ y_i = \mu + g_i + e_i \]

The \( g_i \) are assumed to be randomly drawn from some population with mean = 0 and variance = \( \sigma_g^2 \). Similarly, the \( e_i \) are assumed to be randomly drawn from a population with mean = 0 and variance = \( \sigma_e^2 \). The \( g_i \) and \( e_i \) are assumed to be uncorrelated. If \( \mu \) is known, the least squares estimate of \( g_i \) is merely \( y_i - \mu \). Consequently, the expected value of \( \hat{g}_i \) is \( g_i + e_i \). Then \[ \operatorname{E}[\hat{g}_i] = \sigma_g^2, \quad \text{and} \quad \operatorname{E}[^2] = \sigma_g^2 + \sigma_e^2. \] Consequently, the regression of \( g_i \) on \( \hat{g}_i \) is

\[ \text{Heritability} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} \]

In the case of estimating most probable producing ability from \( n \) records the formula derived by Lush (1945) is

\[ P = \mu + \frac{nr}{1+(n-1)r} (\bar{C} - \mu), \]

where \( P \) denotes most probable producing ability, \( \mu \) denotes herd average, \( n \) denotes number of records, \( r \) denotes repeatability, and \( \bar{C} \) denotes the average of records of the same animal. \( r \) is an intraclass correlation derived by estimating \[ \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}, \] where \( \sigma_a^2 \) is the variance among animals and \( \sigma_e^2 \) is the error variance.

In order to show that the two methods give identical results, the mathematical model, \( y_{ij} = \mu + a_i + e_{ij} \), will be utilized. \( y_{ij} \) is the \( j \)th record of the \( i \)th animal, \( \mu \) is the herd average, \( a_i \) is the real producing ability of the \( i \)th animal, and \( e_{ij} \) is the error associated with the \( ij \)th record. There are \( n_i \) records on the \( i \)th animal. \( E(a_i)^2 = \sigma_a^2 \) and \( E(e_{ij})^2 = \sigma_e^2 \). \( \mu \) is assumed to be known.
The least squares estimate, \( \hat{a}_i \), of \( a_i \) is \( \frac{1}{n_i} \sum_j e_{ij} \). Consequently, \( E(a_i \hat{a}_i) = \sigma_a^2 \), \( E(\hat{a}_i)^2 = \sigma_a^2 + \frac{\sigma_e^2}{n_i} \), and

\[
\hat{a}_i = \frac{\sigma_a^2}{\sigma_a^2 + \frac{\sigma_e^2}{n_i}} (\bar{y}_{ij} - \mu) = \frac{n_i r}{1 + (n_i - 1) r} (\bar{y}_{ij} - \mu), \quad \text{since} \quad r = \frac{\sigma_e^2}{\sigma_a^2 + \sigma_e^2}.
\]

2. **Computational procedure for adjusting least squares estimates**

Having shown that the adjusted least squares estimate is the logical estimate, the problem of obtaining estimates reduces to one of writing the least squares estimate in terms of the population parameters. This can be accomplished if the inverse matrix of the least squares equations is available since this matrix makes it possible to express each of the estimates as a linear function of the sample observations. Let us assume that we are concerned with the linear model \( y_{ijk} = \mu + a_i + b_j + c_{ijk} + e_{ijk} \). Then \( a_i = \sigma_a^2 \sum_i a_i y_{...i} - \sum_i a_i b_j y_{...j} - \sum_i a_i c_{ijk} y_{...j} - \sum_i a_i e_{ijk} \).

In terms of the population parameters \( \hat{a}_i = \frac{p - 1}{p} a_i - \frac{1}{p} \sum_{i} a_i + \sum_{ij} a_{ij} \), \( a_{ij} = \sum_{ijk} \hat{a}_{ijk} e_{ijk} \), where \( a_{ij} \) and \( a_{ijk} \) are coefficients of \( a_{ij} \) and \( a_{ijk} \) peculiar to \( a_i \).

It can be seen that the least squares estimate of \( a_i \) has in it nothing of \( \mu, b_j, \) or \( c \), but does have some of all the other \( a_i \), the \( a_{ij} \), and errors. The following general characteristics with respect to least squares estimates can be stated:

(1) \( E \hat{\mu} \) includes \( \mu \) and \( \frac{1}{p} \) of each of the \( a_i \), \( \frac{1}{q} \) of each of the \( b_j \), and \( \frac{1}{pq} \) of each of the \( a_{ij} \).
(2) $E \hat{a}_s$ always includes $\frac{p-1}{p} a_s$ and $-\frac{1}{p}$ times each of the other $a_i$.

It includes none of $\mu$, $b_j$, or regressions. It includes interaction terms of all order in which $A$ appears.

(3) $E \hat{c}$, the estimate of regression, has in it none of the other parameters.

All estimates have in them error terms.

Once $\hat{a}_i$ is written in terms of the parameters, $E(a_i \hat{a}_i)$ and $E(\hat{a}_i)^2$ can easily be obtained.

$$E(a_i \hat{a}_i) = \frac{p-1}{p} \sigma_a^2$$

$$E(\hat{a}_s)^2 = \frac{p-1}{p} \sigma_a^2 + \sum_{ij} \lambda_{ij}^2 \sigma_a b^2 + a^s a^s \sigma_e^2$$

Then $b(a_i \hat{a}_i) = \frac{\sigma_a^2}{\frac{p-1}{p} \sigma_a^2 + \sum_{ij} \lambda_{ij}^2 \sigma_a b^2 + a^s a^s \sigma_e^2}$

The above derivation assumes that $\sigma_a^2$, $\sigma_{ab}^2$, and $\sigma_e^2$ are known. This unfortunately is never the case. In many practical problems, however, good estimates of the variances are available from previous studies, and these estimates can be put into the formula. If no previous information is available, it is necessary to obtain estimates from the data on which the least squares analysis is carried out. Two methods for obtaining such estimates are presented in Section II F.

E. Estimation of a Parameter Assumed to be a Linear Function of Randomly Drawn Effects

A problem basic to the planning of any breeding program is the estimation of breeding values of individuals from measurements on the individual
and on its relatives. Ordinarily, simple means are involved, but complex least squares estimates might be available. The question arises as to how to put these estimates together in a linear function to give the best possible estimate of breeding value. Both the problem and the solution to it can be put into a very general mathematical statement.

Let \( a_i \) be sets of parameters, with \( i \) denoting the set. It is assumed that the expectation of \( a_i^2 = \sigma_i^2 \), and the expectation of \( a_i a_j = \sigma_{ij} \). The true breeding value, \( T \), is assumed to be a linear function of the \( a_i \) thus, \( T = \sum_i k_i a_i \). The \( a_i \) are, however, unknown, but there are available least squares estimates, \( \hat{a}_i \), of them. The problem is to develop an index \( I \) of the form \( I = \sum_i b_i \hat{a}_i \) which will give the best possible estimate of \( T \).

Assuming that \( \sigma_i^2 \) and \( \sigma_{ij} \) are known, it appears that the maximum likelihood estimate of \( T \) corrected for the \( \frac{p-1}{p} \) bias is \( \sum_i b_i \hat{a}_i \), where the \( b_i \) are the solution to the set of equations below.

\[
\begin{array}{cccc}
 b_1 & b_2 & \cdots & b_p \\
 \vdots & \vdots & \ddots & \vdots \\
 \sigma_i^2 & \sum_i k_i \hat{a}_i a_1 & \vdots & \vdots \\
 \sum_i k_i \hat{a}_i a_2 & \sigma_{ij} & \sum_i k_i \hat{a}_i a_2 & \vdots \\
 \vdots & \vdots & \ddots & \vdots \\
 \sum_i k_i \hat{a}_i a_p & \vdots & \vdots & \sigma_{ij} \\
 \end{array}
\]
The only facts needed to set up these equations are statements of the \( \hat{\beta}_4 \) as linear functions of the parameters. The method for doing this was described in Section II D 2. This method of estimation is identical with Hazen's (1943) selection index except that the independent variables in his study were measurable characteristics rather than least squares estimates.

F. Estimation of Components of Variance in the Non-Orthogonal Case

The problem of estimation of variance components is in an unsatisfactory state at present since there is little but intuition to serve as a guide in choosing a "best" method of estimation. Crump (1947) has recently discussed this problem in detail and has presented maximum likelihood solutions for a few simple classifications. A great number of ways could be devised for obtaining unbiased estimates, but very little is known about the sampling errors of estimates of variance components. Two methods will be described below. Method I involves computing sums of squares as though the data were orthogonal and one were going to do an analysis of variance. Then expected values are taken of these sums of squares, the expected values are equated to the sample values, and the resulting equations are solved simultaneously to obtain estimates of the variance components. This method enables quick estimates to be made from complicated multiple classifications, either nested or factorial or a combination of the two. In this method all of the effects in the linear mathematical model are assumed to be randomly drawn from certain populations, and the expected values of all cross products are assumed equal to zero. This may not make good sense in some problems.
Method II involves taking expected values of sums of squares calculated by least squares, equating the expectations to the sample values, and solving the equations simultaneously for the estimates of the variances. This method is relatively laborious since it requires inversion of certain matrices, but it has the advantage of permitting the assumption that one or more of the sets of effects are not randomly drawn, and of requiring no assumption about the expected values of the cross products among such effects. Furthermore, at least intuitively, the second method would seem to give better estimates than the first method in badly balanced experiments, since the expectation of the sums of squares for a particular set is freed of everything except the variance being estimated, higher order interactions involving the same letter, and \( \sigma_e^2 \).

For example, in a three-way classification, in method I the expectation of the A sum of squares is \( k_a \sigma_a^2 + k_b \sigma_b^2 + k_c \sigma_c^2 + k_{ab} \sigma_{ab}^2 + k_{ac} \sigma_{ac}^2 + k_{bc} \sigma_{bc}^2 + k_{abc} \sigma_{abc}^2 + k_e \sigma_e^2 \). Only in the orthogonal case does \( k_b = k_c = k_{bc} = 0 \). In method II the expectation of the A sum of squares is \( k'_a \sigma_a^2 + k'_{ab} \sigma_{ab}^2 + k'_{ac} \sigma_{ac}^2 + k'_{abc} \sigma_{abc}^2 + k'_e \sigma_e^2 \).

1. Estimates from a conventional analysis of variance

The estimates arising from a conventional analysis of variance will be illustrated by a two-way factorial classification with unequal subclass numbers. The model is

\[
\begin{align*}
Y_{ijk} & = \mu + a_i + b_j + ab_{ij} + e_{ijk} \\
E(a_i) & = E(b_j) = E(ab_{ij}) = E(e_{ijk}) = 0 \\
E(a_i)^2 & = \sigma_a^2 \\
E(b_j)^2 & = \sigma_b^2 \\
E(ab_{ij})^2 & = \sigma_{ab}^2
\end{align*}
\]
\[ E(e_{ijk})^2 = \sigma_e^2 \]
\[ E(\text{all cross products}) = 0 \]
\[ i = 1, \ldots, p \]
\[ j = 1, \ldots, q \]
\[ k = 1, \ldots, n_{ij} \]

The following sums of squares are calculated:

\[ T = \sum_{ijk} y_{ijk}^2 \]
\[ A = \sum_{i} \frac{1}{n_i} y_{i..}^2 \]
\[ B = \sum_{j} \frac{1}{n_{ij}} y_{..j}^2 \]
\[ AB = \sum_{ij} \frac{1}{n_{ij}} y_{ij}^2 \]
\[ \text{Corr.} = \frac{1}{n_{..}} y_{..}^2 \]

The expectation of these sums of squares are

\[ E(T) = n_{..} \left( \mu^2 + \sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 + \sigma_e^2 \right) \]
\[ E(AB) = n_{..} \left( \mu^2 + \sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 \right) + (pq - m) \sigma_e^2, \text{ where } m = \]
\[ \text{number of missing subclasses} \]
\[ E(A) = n_{..} \mu^2 + n_{..} \sigma_a^2 + \frac{\sum_{i} \frac{n_{i..}}{n_i} \sigma_{b^2} + \sigma_{ab}^2}{n_i} \]
\[ E(B) = n_{..} \mu^2 + n_{..} \sigma_b^2 + \frac{\sum_{j} \frac{n_{..j}}{n_{ij}} \sigma_{a^2} + \sigma_{ab}^2}{n_{ij}} \]
\[ E(\text{corr.}) = n_{..} \mu^2 + \frac{1}{n_{..}} \left[ \sum_{i} \frac{n_{i..}}{n_i} \sigma_a^2 + \sum_{j} \frac{n_{..j}}{n_{ij}} \sigma_b^2 + \sum_{ij} \frac{n_{ij}}{n_{ij}} \sigma_{ab}^2 \right] \]
\[ + \sigma_e^2. \]

In the usual analysis of variance one would obtain various sums of squares as follows:

\[ A \text{ S.S.} = A - \text{Corr.} \]
\[ A \times B \text{ S.S.} = AB - A - B + \text{Corr.} \]
\[ B \text{ S.S.} = B - \text{Corr.} \]
\[ \text{Error S.S.} = T - AB \]
The expected values of these sums of squares can easily be obtained from the previously calculated ones. For example, \( E(A \text{ S.S.}) = E(A - \text{Corr.}) \) = \( E(A) - E(\text{Corr.}) \).

The following equations are now available for estimating the variances:

\[
\begin{align*}
    k_{11} \sigma_a^2 + k_{12} \sigma_b^2 + k_{13} \sigma_{ab}^2 + k_{14} \sigma_e^2 &= A \text{ S.S.} \\
    k_{21} \sigma_a^2 + k_{22} \sigma_b^2 + k_{23} \sigma_{ab}^2 + k_{24} \sigma_e^2 &= B \text{ S.S.} \\
    k_{31} \sigma_a^2 + k_{32} \sigma_b^2 + k_{33} \sigma_{ab}^2 + k_{34} \sigma_e^2 &= A \times B \text{ S.S.} \\
    k_{44} \sigma_e^2 &= \text{Error S.S.}
\end{align*}
\]

The solution to these equations yields mean unbiased estimates of the variances. That is, \( E(\hat{\sigma}_1^2) = \sigma_1^2 \).

The extension of this method to a more than two-way classification is straightforward and involves no new principles.

2. **Estimates from least squares sum of squares**

Since method II involves taking expected values of reductions in sums of squares due to fitting certain parameters, it will be necessary first to present a general mathematical statement of \( E R(\quad) \). Assume that there are parameters \( a_1, a_2, \ldots, a_p \) with associated sums \( y_1, y_2, \ldots, y_p \). The \( a \)'s may include several different sets of parameters. For example, \( a_1, a_2, \ldots, a_q \) may represent \( b_1, b_2, \ldots, b_q \); \( a_{q+1}, \ldots, a_r \) may represent \( c_1, \ldots, c_{r-q} \); etc.

The inverse of the matrix of coefficients of the least squares equations is,
Then \( \hat{a}_j = \sum_i a_i a_j y_i \), and \( R(a) = \sum_j \hat{a}_j y_j = \sum_i a_i a_j y_i y_j \). Therefore,

\[
E[R(a)] = E \left[ \sum_{ij} a_i a_j y_i y_j \right] = \sum_{ij} E[a_i a_j y_i y_j].
\]

That is, the expected value of the reduction in sum of squares is obtained by inverting the matrix of coefficients in the least squares equations, multiplying each element of the inverse by the expected value of the products of the sums associated with the row and column of the element, and then summing all of such products.

Since \( E(y_i y_j) \) is very easily obtained, the computational labor in this method arises largely in inverting the matrix. But since the inverse matrix is needed anyway for obtaining sampling errors of parameters estimated by least squares, for computing least squares estimates when several different types of observations are taken on the same material, and for obtaining \( b(a_i a_j) \), the estimation of variance components from least squares sums of squares is often little extra labor. In addition, general rules can be stated which make it possible to write down certain expectations without resorting to the invert matrix. The first of these rules is that
it is a characteristic of least squares sums of squares that the expectation of the reduction in sums of squares due to fitting several sets of parameters minus the reduction due to fitting another set of parameters has in it no squares nor cross products involving the parameters common to both reductions. For example, \( E \{ \hat{R}(\mu, a, b, c) - \hat{R}(\mu, a, b) \} \) has in it no squares nor cross products involving \( \mu, a, \) or \( b \). This is the reason why the least squares method makes possible the assumption that certain of the effects are not randomly drawn from some population of effects. For example, in the above illustration, it might be assumed that the \( a_1 \) are arbitrarily chosen and the \( b_j \) and \( c_k \) are randomly drawn from two different populations. Then the expectation of \( \hat{R}(\mu, a, b, c) - \hat{R}(\mu, a, b) \) is equal to \( k_1 \sigma^2 + k_2 \sigma_e^2 \), assuming that interactions are considered to be non-existent.

The second useful rule for writing the expectation of least squares sums of squares is that the expectation of the coefficients of the variance pertaining to a given set of parameters fitted in a particular reduction is equal to the sum of the squared coefficients of the members of the set appearing in the mathematical model. For example,

1. if \( y = \mu + a_1 + \ldots \),
   \[
   E \{ \hat{R}(\mu, a, \ldots) \} = n \mu^2 + n \sigma_a^2 + \ldots ;
   \]
2. if \( y = \mu + \varepsilon_1 + \varepsilon_j + \ldots \),
   \[
   E \{ \hat{R}(\mu, \varepsilon, \ldots) \} = n \mu^2 + 2n \sigma_\varepsilon^2 + \ldots ;
   \]
3. if \( y = \mu + 2a_1 + \ldots \),
   \[
   E \{ \hat{R}(\mu, a, \ldots) \} = n \mu^2 + 4n \sigma_a^2 + \ldots .
   \]

In all cases \( n \) = the total number of observations.

The third rule is that the coefficient of \( \sigma_e^2 \) in any reduction is equal to the number of independent parameters estimated. For example, in a \( p \times q \)
For purposes of illustrating the method the mathematical model, \( y_{ijk} = \mu + a_1 + b_i + c_j + d_{ij} + a_{ijk}, \) is utilized. \( \mu, a_i, \) and \( b_i \) are constants while \( c_j, d_{ij}, \) and \( a_{ijk} \) are randomly drawn from populations with mean = 0 and variances \( \sigma^2_c, \sigma^2_d, \sigma^2_{cd}, \) and \( \sigma^2_e. \) It is assumed further that the cross products among these latter effects are equal in all cases to zero. Estimates of \( \sigma^2_c, \sigma^2_d, \sigma^2_{cd}, \) and \( \sigma^2_e \) are to be obtained by taking expected values of various \( R( \ ) \)'s and equating to sample reductions in sums of squares. Utilizing the general rules presented above, the expected values of pertinent reductions are as follows: \( E \{ R(\mu, a, b, c, d, cd) \} = S + n \left( \sigma^2_c + \sigma^2_d + \sigma^2_{cd} \right) + (p + q + rs - m - 2) \sigma^2_e, \) where \( m = \) the number of missing subclasses, \( n = \) the total number of observations, and \( S = \) the expectation of the various squares and cross products arising from \( \mu, a_i, \) and \( b_i. \) \( k_1, \ldots, k_5 \) are coefficients which need to be computed.

\[
E \{ R(\mu, a, b, c, d) \} = S + n \left( \sigma^2_c + \sigma^2_d + \sigma^2_{cd} \right) + \frac{k_1 \sigma^2_{cd}}{3} \cdot (p + q + r + s - 3) \sigma^2_e.
\]

\[
E \{ R(\mu, a, b, c) \} = S + n \sigma^2_c + k_2 \sigma^2_d + k_3 \sigma^2_{cd} + (p + q + r - 2) \sigma^2_e.
\]

\[
E \{ R(\mu, a, b, d) \} = S + k_4 \sigma^2_c + n \sigma^2_d + k_5 \sigma^2_{cd} + (p + q + s - 2) \sigma^2_e.
\]

\[
E \{ R(T) \} = S + n \left( \sigma^2_c + \sigma^2_d + \sigma^2_{cd} + \sigma^2_e \right).
\]

Let \( C = R(\mu, a, b, c, d) - R(\mu, a, b, d) \)
\( D = R(\mu, a, b, c, d) - R(\mu, a, b, c) \)
\( C \times D = R(\mu, a, b, c, d, cd) - R(\mu, a, b, c, d) \)
\( Error = R(T) - R(\mu, a, b, c, d, cd). \)
If the expected values of C, D, C x D, and error are equated to the corresponding calculated sums of squares, the following set of equations is obtained, the solution to which gives mean unbiased estimates of the variance components:

<table>
<thead>
<tr>
<th>$\delta^2_c$</th>
<th>$\delta^2_d$</th>
<th>$\delta^2_{cd}$</th>
<th>$\delta^2_e$</th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>n - k₄</td>
<td>0</td>
<td>$k_1 - k_5$</td>
<td>r - 1</td>
<td>C</td>
</tr>
<tr>
<td>0</td>
<td>n - k₂</td>
<td>$k_1 - k_3$</td>
<td>s - 1</td>
<td>D</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>n - k₄</td>
<td>$(r-1)(s-1) - m$</td>
<td>C x D</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$n - p - q - rs$</td>
<td>Error</td>
</tr>
</tbody>
</table>

The problem therefore reduces to one of finding the values of $k_2$, $k_3$, $k_5$. As previously described, this involves inverting the appropriate matrix of coefficients of least squares equations and then multiplying each element of the matrix by the expected value of the corresponding product between associated sums. It is not, of course, necessary to compute the complete expectations, the only computations required being the coefficients of the particular variances involved.

It is now apparent that the least squares method provides a computational procedure not only for correcting data for extraneous sources of variation preparatory to estimating a set of population values, but it also provides a method for obtaining sums of squares freed of the extraneous sources of variation. Estimates of population variances can then be obtained from these "corrected" sums of squares.
G. Application of the Methods to the Analysis of Single Cross Data

The mathematical model used for analysis of the single cross data was

\[ Y_{ijkl} = \mu + e_i + g_j + m_k + s_{ij} + r_{ij} + a_{jk} + e_{ijkl} \]

The assumptions made concerning the components of the model were described in Section II A. A least squares analysis is first carried out in order to make tests of hypotheses and in order to correct the class mean for extraneous sources of variation. The least squares equations are obtained by differentiating the error sum of squares, \[ \sum(y_{ijkl} - \mu - e_i - g_j - m_k - s_{ij} - r_{ij} - a_{jk})^2 \]
partially with respect to each of the parameters and then setting each derivative equal to zero. The following equations arise:

\[ \begin{align*}
\mu: & \quad n \ldots \mu + \sum_{i=j} (n_{i..} + n_{.j..}) g_i + \sum_{j\neq i} n_{.j.} m_j + \sum_{ij} (n_{ij.} + n_{.ij}) \\
& \quad \quad s_{ij} + \sum_{ij} n_{ij.} r_{ij} + \sum_k n_{..k} a_k = y_{..} \\
g_w: & \quad (n_{w..} + n_{..w}) (\mu + g_w) + \sum_{i=j=w} (n_{w..} + n_{i..w}) (g_i + s_{iw}) + n_{..w} \\
& \quad m_w + \sum_{i\neq w} n_{w..} (m_j + r_{wj}) + \sum_{i\neq w} n_{iw} r_{iw} + \sum_k (n_{..k} + n_{.wk}) a_k \\
& \quad = y_{..} + y_{w..} \\
m_w: & \quad n_{w..} (\mu + g_w + m_w) + \sum_{i\neq w} n_{iw} (g_i + s_{iw} + r_{iw}) + \sum_k n_{..k} a_k \\
& \quad = y_{..} \\
s_vw: & \quad (n_{vw..} + n_{..vw}) (\mu + g_v + g_w + s_{vw}) + n_{vw..} (m_v + r_{vw}) + n_{..vw} \\
& \quad (m_v + r_{vw}) + \sum_k (n_{vwk} + n_{wvk}) a_k = y_{..} + y_{vw..} + y_{..v} \\
r_vw: & \quad n_{vw..} (\mu + g_v + g_w + m_v + s_{vw} + r_{vw}) + \sum_k n_{vwk} a_k = y_{vwk}.
\end{align*} \]
\[ a_{i} = n_{i} \cdot (\mu + e_{i}) + \sum_{j} \left( n_{i,j} + n_{j,i} \right) s_{ij} + \sum_{j} n_{ij} m_{j} + \sum_{ij} \]

The parameter followed by a colon denotes the equation obtained when \( \sum_{ijkl} e_{ijkl}^2 \) is differentiated with respect to that parameter.

1. Tests of hypotheses

The problem of order of making tests arises immediately. The following procedure appears to be a logical one for this problem. First the hypothesis that \( r_{ij} = 0 \) is tested. If this hypothesis is accepted, the next step follows logically. If the hypothesis is rejected, however, subsequent tests made under the assumption that \( r_{ij} \neq 0 \) are almost impossibly difficult when the number of lines is at all appreciable. In what is to follow it will be assumed, therefore, that the hypothesis that \( r_{ij} = 0 \) is accepted. Next the hypothesis is tested that \( s_{ij} = 0 \). If this hypothesis is accepted, the hypothesis that \( m_{j} = 0 \) is then tested under the assumption that \( s_{ij} = 0 \). If the hypothesis that \( s_{ij} = 0 \) is rejected, the hypothesis that \( m_{j} = 0 \) can be tested, assuming \( s_{ij} \neq 0 \). The order of testing \( s_{ij} \) and \( m_{j} \) might well be reversed. Finally, the hypothesis that \( g_{i} = 0 \) is tested, the method of doing this depending upon whether previous tests have led to acceptance or rejection of the hypothesis that \( m_{j} = 0 \) and the hypothesis that \( s_{ij} = 0 \). If it is assumed that \( s_{ij} \neq 0 \), the test of the hypothesis that \( g_{i} = 0 \) is very difficult since the coefficients of \( s_{ij} \) cannot be absorbed.

The methods for making tests under different assumptions will now be described. \( R(\ldots) \) as before denotes the reduction in sum of squares due to fitting a particular set of parameters. It is assumed that there are
p lines, q crosses ignoring which way the crosses were made, t reciprocal crosses, v A classes, and a total of n observations. Each of the p lines appears at least once as line of sire and as line of dam. \( R(T) \) denotes \( \sum y_{ijkl}^2 \).

(a) Analyses of variance. The analyses of variance for several different tests are as follows:

1. The hypothesis that \( r_{ij} = 0 \)
   \[ \frac{df}{S.S.} \]
   Among \( r \) \( t-p-q+1 \) \[ R(\mu, g, m, a) - R(\mu, g, m, s, r, a) \]
   Error \( n-t-v+1 \) \[ R(T) - R(\mu, g, m, s, r, a) \]

2. The hypothesis that \( s_{ij} = 0 \) assuming \( r_{ij} = 0 \)
   \[ \frac{df}{S.S.} \]
   Among \( s \) \( q-p \) \[ R(\mu, g, m, s, a) - R(\mu, g, m, a) \]
   Error \( n-p-q-v+2 \) \[ R(T) - R(\mu, g, m, s, a) \]

3. The hypothesis that \( s_{ij} = 0 \) assuming \( m_{ij} = r_{ij} = 0 \)
   \[ \frac{df}{S.S.} \]
   Among \( s \) \( q-p \) \[ R(\mu, g, m, s, a) - R(\mu, g, a) \]
   Error \( n-q-v+1 \) \[ R(T) - R(\mu, g, a) \]

4. The hypothesis that \( m_{ij} = 0 \) assuming \( r_{ij} = 0 \)
   \[ \frac{df}{S.S.} \]
   Among \( m \) \( p-1 \) \[ R(\mu, g, m, s, a) - R(\mu, g, a) \]
   Error \( n-p-q-v+2 \) \[ R(T) - R(\mu, g, m, s, a) \]

5. The hypothesis that \( m_{ij} = 0 \) assuming \( s_{ij} = r_{ij} = 0 \)
   \[ \frac{df}{S.S.} \]
   Among \( m \) \( p-1 \) \[ R(\mu, g, m, a) - R(\mu, g, a) \]
   Error \( n-2p-v+2 \) \[ R(T) - R(\mu, g, m, a) \]

6. The hypothesis that \( m_{ij} = 0 \) assuming \( s_{ij} = s_{ij} = r_{ij} = 0 \)
   \[ \frac{df}{S.S.} \]
   Among \( m \) \( p-1 \) \[ R(\mu, m, a) - R(\mu, a) \]
   Error \( n-p-v+1 \) \[ R(T) - R(\mu, m, a) \]

7. The hypothesis that \( g_{ij} = 0 \) assuming \( s_{ij} = r_{ij} = 0 \)
   \[ \frac{df}{S.S.} \]
   Among \( g \) \( p-1 \) \[ R(\mu, g, m, a) - R(\mu, m, a) \]
   Error \( n-2p-v+2 \) \[ R(T) - R(\mu, g, m, a) \]

8. The hypothesis that \( g_{ij} = 0 \) assuming \( m_{ij} = s_{ij} = r_{ij} = 0 \)
   \[ \frac{df}{S.S.} \]
   Among \( g \) \( p-1 \) \[ R(\mu, g, a) - R(\mu, a) \]
   Error \( n-p-v+1 \) \[ R(T) - R(\mu, g, a) \]
b. Computation of reductions in sums of squares. Computational procedures for obtaining the various $R$'s are described below.

1. $R(\mu, g, m, s, r, a)$. The leading equation for $r_{ij}$ is $n_{ij}$.

$$ (\mu + g_i + g_j + m_i + s_{ij} + r_{ij}) \sum_k n_{ijk} a_k = y_{ij}^{..} $$

Consequently the computation is greatly facilitated by letting

$$ \tilde{r}_{ij} = \mu + g_i + g_j + m_i + s_{ij} + r_{ij}^{..} $$

Then since the $\mu$, $g_i$, $m_j$, and $s_{ij}$ equations are each the sum of a particular set of the $r_{ij}$ equations, they can be deleted leaving a set of equations involving only $r_{ij}$ and $a_k$, thus

$$ \tilde{r}_{uv} = \mu + g_i + g_j + m_i + s_{ij} + r_{ij}^{..} $$

$$ a_k = \sum_{ij} n_{ijk} \tilde{r}_{ij} + y_{ij}^{..} k $$

But since $\tilde{r}_{ij} = \frac{1}{n_{ij}} (y_{ij}^{..} - \sum_k n_{ijk} a_k)$, the $\tilde{r}_{ij}$ coefficients can be absorbed leaving a set of equations involving only $a_k$ for solution. The values obtained for $a_k$ can then be substituted in the equations involving $\tilde{r}_{ij}$ and the complete solution obtained. Finally, $R(\mu, g, m, s, r, a)$

$$ = \sum_{ij} \tilde{r}_{ij} y_{ij}^{..} + \sum_k a_k y_{..}k $$

2. $R(\mu, g, m, s, a)$. The leading equation for $s_{ij}$ is $(n_{ij} + n_{ji})(\mu + g_i + g_j + s_{ij}) + n_{ij}. m_j + n_{ji}. m_i + \sum_{k} (n_{ijk} + n_{ikj}) a_k = y_{ij}^{..} + y_{ji}^{..}$

Let $\tilde{s}_{ij} = \mu + g_i + g_j + s_{ij}^{..}$. Then since the $\mu$ and $g_i$ equations are each sums of certain of the $s_{ij}$ equations, they can be deleted leaving
\[
m_u: \ n \cdot u \cdot m_u + \sum_i n_{iu} \cdot \hat{s}_{iu} + \sum_k n_{uk} \cdot a_k = y_u...
\]
\[
\tilde{u}_v: \ n \cdot u_v \cdot m_v + \sum_v n_{vu} \cdot m_u + (n_{uv} + n_{vu}) \cdot \hat{s}_{uv} + \sum_k (n_{uvk} + n_{vuk}) \cdot a_k = y_u + y_v...
\]
\[
a_k: \ \sum_j n \cdot jk \cdot m_j + \sum_{ij} (n_{ijk} + n_{jik}) \cdot \hat{s}_{ij} + n \cdot k \cdot a_k = y \cdot k.
\]

But since \( \hat{s}_{uv} = \frac{1}{n_{uv} + n_{vu}} \left[ y_{uv} + y_{vu} - n_{uv} \cdot m_v - n_{vu} \cdot m_u - \sum_k (n_{uvk} + n_{vuk}) \cdot a_k \right] \), the equations can be reduced to ones involving only \( m_j \) and \( a_k \). After eliminating the two dependent parameters the equations can be solved, and \( R(\mu, g, m, s, a) = \sum_j \hat{y}_{ij} \cdot y_{ij} + \sum_{ij} \hat{s}_{ij} \cdot (y_{ij} + y_{ji}) + \sum_k \hat{a}_k \cdot y_{k} \).

(3) \( R(\mu, g, s, a) \). The procedure for computing this reduction is exactly the same as for \( R(\mu, g, m, s, a) \) except that the \( m_j \) are deleted from all equations and the \( m_j \) equations are deleted.

(4) \( R(\mu, g, m, a). \) Let \( d_j = \mu + g_j + m_j \). The least squares equations then are
\[
g_u: \ n_u \cdot g_u + \sum_{j \neq u} n_{uj} \cdot d_j + \sum_k n_{uk} \cdot a_k = y_u...
\]
\[
d_u: \ \sum_{j \neq u} n_{iu} \cdot g_u + n_u \cdot d_u + \sum_k n_{uk} \cdot a_k = y_u...
\]
\[
a_k: \ \sum_i n \cdot i.k \cdot g_i + \sum_j n \cdot jk \cdot d_j + n \cdot k \cdot a_k = y \cdot k.
\]

The \( d_j \) coefficients can be absorbed since \( d_j = \frac{1}{n_{.j}} \cdot (y_{.j} + \sum_{i} n_{ij} \cdot g_i - \sum_k n_{jk} \cdot a_k) \). The equations are thereby reduced to ones containing only \( g_i \) and \( a_k \). Then \( R(\mu, g, m, a) \)
(5) \( R(\mu, \varepsilon, \sigma) \). Let \( \hat{\alpha}_k = \mu + a_k \). The \( \mu \) equation can be deleted since it is merely the sum of the \( a_k \) equations. Then the following set needs to be solved.

\[
E_u: (n_{u..} + n_{u.}) \varepsilon_u + \sum_{i \neq j} (n_{iju.} + n_{uj.}) \varepsilon_i + \sum_k (n_{u.k} + n_{u.}) \hat{a}_k = y_{u..} + y_{u..}
\]

\[
\hat{a}_k: \sum_{i \neq j} (n_{i.k} + n_{j.k}) \varepsilon_i + n_{..k} \hat{a}_k = y_{..k}
\]

Since \( \hat{a}_k = \frac{1}{n_{..k}} \left[ y_{..k} - \sum_{i \neq j} (n_{i.k} + n_{j.k}) \varepsilon_i \right] \), the equations can be reduced to ones containing only \( \varepsilon_i \), and

\[
R(\mu, \varepsilon, \sigma) = \sum_i \varepsilon_i (y_{i..} + y_{j..}) + \sum_k \hat{a}_k y_{..k}
\]

(6) \( R(\mu, m, a) \). Let \( \tilde{m}_j = \mu + m_j \). Then the equations are

\[
\tilde{m}_j: n_{.j.} \tilde{m}_j + \sum_k n_{.jk} a_k = y_{.j..}
\]

\[
a_k: \sum_j n_{.jk} \tilde{m}_j + n_{..k} a_k = y_{..k}
\]

Since \( \tilde{m}_j = \frac{1}{n_{.j.}} (y_{.j..} - \sum_k n_{.jk} a_k) \), the equations can be reduced to a set containing only \( a_k \). Then \( R(\mu, m, a) = \sum_i \tilde{a}_i y_{i..} + \sum_k \hat{a}_k y_{..k} \).

(7) \( R(\mu, a) \). Let \( \hat{\alpha}_k = \mu + a_k \). Then the equations are

\[
\hat{\alpha}_k: n_{..k} \hat{\alpha}_k = y_{..k}
\]

Consequently, \( \hat{\alpha}_k = \frac{1}{n_{..k}} y_{..k} \).

\[
R(\mu, a) = \sum_k \hat{a}_k y_{..k} = \sum_k \frac{1}{n_{..k}} y_{..k}^2
\]

2. Least squares estimates

Computation of certain of the \( R(\_\_\_\_\_\_\_) \)'s in connection with tests of
hypotheses were effected without estimating the individual parameters since
\( R(\quad) \) can be computed more easily by considering some joint parameter.
Parameter estimates can, however, be extracted from these joint parameter
estimates if they are needed. Furthermore, it will be noted that several
different estimates of parameters are available depending upon what as-
sumptions are made. The choice of estimate should depend upon the re-
sults of the tests of hypotheses. For example, if it should be found that
the hypotheses that \( r_{ij} = 0 \) and that \( s_{ij} = 0 \) are both accepted, but that
the hypotheses that \( m_j = 0 \) and that \( g_i = 0 \) are both rejected, the esti-
mates of \( m_j \) and \( g_i \) should be obtained from the estimates arising from
\( R(\mu, g, m, s, a) \). The estimates arising from \( R(\mu, g, m, s, r, a) \) are also unbiased, but the sampling errors are likely to be
higher, particularly if the design is far from a balanced one. In the
case of a completely balanced design the sampling errors are entirely a
function of \( \frac{\sigma^2}{n} \) and the number of observations, since \( g_i \) and \( m_j \) are or-
thogonal to \( s_{ij} \) and \( r_{ij} \) (Section II 11 a).

Methods for extracting parameter estimates from the computations of
the various \( R(\quad) \)'s will now be presented.

a. Estimates of \( \mu, g_i, m_j, s_{ij}, \) and \( r_{ij} \) from \( R(\mu, g, m, s, r, a) \).

In II 11 b estimates were obtained of \( \tilde{r}_{ab} \), where \( \tilde{r}_{ab} = \mu + g_a + g_b + m_a + m_b + s_{ab} + r_{ab} \). Since \( r_{ab} + r_{ba} = 0 \) and since \( r_{ab} = 0 \) if \( \tilde{n}_{ab} = 0 \), the estimates
of \( \mu, g_i, m_j, \) and \( s_{ij} \) can be obtained from the following equations:

\[
\begin{align*}
\mu &= n_{..}^{-1} \sum_{i,j} (n_{i.} + n_{.j}) \bar{g}_i + \sum_j n_{.j} \bar{m}_j + \sum_i (n_{ij} + n_{ji}) \bar{s}_{ij} \\
&= \tilde{r}_{..}
\end{align*}
\]

\[
\begin{align*}
g_a &= (n_{..} + n_{.a}) (\mu + g_a) + \sum_i (n_{aj} + n_{ia}) (g_i + s_{ai}) + n_{.a} m_a + \\
&\quad \sum_j n_{aj} m_j \quad \tilde{r}_{.a} + \tilde{r}_a
\end{align*}
\]
The best way to solve these equations is the method described in Section II G 1 b (2). That is, the coefficients of $\hat{s}_{ab} = \hat{\mu} + \hat{\varepsilon}_a + \hat{\varepsilon}_b + s_{ab}$ are absorbed into the $m_j$ coefficients. The $\hat{\mu}_j$ arising from this solution are the least squares estimates of $m_j$ under the assumption that $\varepsilon_1$, $s_{ij}$, $r_{ij}$, $m_j \neq 0$.

The estimates of $\mu$, $\varepsilon_1$, and $s_{ij}$ are then obtained from the estimates of $\hat{s}_{ij}$ just computed. Utilizing the fact that $\sum_i s_{ij} = 0$, the following equations need to be solved:

$$\mu: \quad \bar{n} \cdot \mu + \sum_i n_i \varepsilon_i = \hat{s}_{..}$$

$$\varepsilon_a: \quad \bar{n} \cdot (\varepsilon_a + A) + \sum_i n_i \varepsilon_i = \hat{s}_a.$$

The solution to these equations gives $\hat{\mu}$ and $\hat{\varepsilon}_1$. Then $\hat{s}_{ij} = \hat{s}_{ij} - \hat{\mu} - \hat{\varepsilon}_1 - \hat{\varepsilon}_j$, and $\hat{r}_{ij} = \hat{r}_{ij} - \hat{\mu} - \hat{\varepsilon}_1 - \hat{\varepsilon}_j - \hat{m}_j - \hat{s}_{ij}$.

b. Estimates of $\mu$, $\varepsilon_1$, $m_j$, and $s_{ij}$ from $R(\mu, \varepsilon, m, s, a)$. In the computation of $R(\mu, \varepsilon, m, s, a)$ estimates of $m_j$ were obtained directly. Estimates also arose of $\hat{s}_{ij}$, where $\hat{s}_{ij} = \hat{\mu} + \hat{\varepsilon}_1 + \hat{\varepsilon}_j + s_{ij} \cdot \hat{\mu}$, $\hat{\varepsilon}_1$, and $\hat{s}_{ij}$ can now be computed from the $\hat{s}_{ij}$ exactly as described in Section II G 2 a.

c. Estimates of $\mu$, $\varepsilon_1$, and $s_{ij}$ from $R(\mu, \varepsilon, s, a)$. $\hat{s}_{ij}$ was estimated in the computation of this reduction in sum of squares. Estimates of
\( \mu, g_i, \) and \( s_{ij} \) are obtained from \( \hat{a}_{ij} \) exactly as was described in Section II C 2 a.

d. Estimates of \( \mu, g_i, \) and \( m_j \) from \( R(\mu, g, m, a) \). Direct estimates of \( g_i \) were obtained in the computation of this reduction. Estimates of \( d_j \) also arose where \( d_j = \mu + g_j + m_j. \)

Since \( \sum_j g_j = \sum_j m_j = 0; \sum_j d_j = p \mu, \hat{\mu} = \frac{1}{p} \sum_j \hat{d}_j, \) and \( \hat{m}_j = \hat{d}_j - \hat{\mu} - \hat{g}_j. \)

e. Estimates of \( \mu \) and \( g_i \) from \( R(\mu, g, a) \). Estimates of \( g_i \) were obtained directly in the computation of this sum of squares.

f. Estimates of \( \mu \) and \( m_j \) from \( R(\mu, m, a) \). Estimates were obtained of \( \hat{m}_j = \mu + m_j. \) Therefore, \( \hat{\mu} = \frac{1}{p} \sum_j \hat{m}_j; \hat{m}_j = \hat{m}_j - \hat{\mu}. \)

3. Sampling errors of the least squares estimates

As was discussed in Section II C 2, different estimates of parameters are obtained depending upon what assumptions are made with respect to the other parameters. Similarly, different sampling errors of estimates are obtained depending upon what assumptions are made. Methods will now be described for obtaining sampling errors under several different assumptions.

a. \( \xi_{ij}, m_j, g_i, r_{ij}, f_{ij} \neq 0. \) The procedure described for obtaining \( R(\mu, g, m, a, r, a) \) suggests the procedure for inverting the matrix of coefficients of the least square equations. That is, let \( \tilde{r}_{ij} = \mu + g_i + g_j + m_j + s_{ij} + r_{ij} \) and then absorb this set of parameters into the coefficients of \( \xi_k, \) invert the resulting matrix involving only \( \xi_k, \) and from that inverse matrix calculate the inverse elements corresponding to the \( \tilde{r}_{ij}. \)

Then the products of these inverse elements by \( \tilde{\sigma}^2 \) yield the variance-covariance matrix of the \( \hat{r}_{ij} \) and \( \hat{\xi}_k. \) To go from there to the variance-covariance matrix of \( \hat{\mu}, \hat{g}_i, \hat{m}_j, \hat{s}_{ij}, \) and \( \hat{r}_{ij} \) involves expressing each of
the estimates as a linear function of the $\hat{r}_{ij}$. As was described in Section II.2, $m_j$ and $\hat{s}_{ij} = \mu + g_i + g_j + s_{ij}$ are estimated from the equations:

$$n_j m_j + \sum_i n_{ij} \hat{s}_{ij} = \sum_i \hat{r}_{ij}$$

$$n_{ij} \mu + n_{ij} m_i + (n_{ij} + n_{ji}) \hat{s}_{ij} = \hat{r}_{ij} + \hat{r}_{ji}$$

If the inverse of this matrix is now taken, $\hat{m}_j$ and $\hat{s}_{ij}$ can be expressed as linear functions of $\hat{r}_{ij}$. Thus,

$$\hat{m}_j = \sum_j m_j u \hat{r}_{ij} + \sum_{ij} m_j g_{ij} (\hat{r}_{ij} + \hat{r}_{ji}) = \sum_{ij} \lambda_{ij} \hat{r}_{ij}$$

Similarly, $\hat{s}_{ij} = \sum_{ij} c_{ij} \hat{r}_{ij}$

$\lambda_{ij}$ and $c_{ij}$ denote coefficients of $\hat{r}_{ij}$ peculiar to $\hat{m}_j$ and $\hat{s}_{ij}$ respectively.

Since $\mu$ and $g_i$ are estimated from the $\hat{s}_{ij}$ by the equations,

$$n_j \mu + \sum_i n_{ij} \hat{s}_{ij} = \sum_{ij} \hat{s}_{ij}$$

$$n_{ij} \mu + n_{ij} \hat{s}_{ij} + \sum_{ij} n_{ij} \hat{g}_j = \sum_{ij} \hat{s}_{ij}$$

$\hat{\mu}$ and $\hat{g}_i$ can be expressed as linear functions of $\hat{s}_{ij}$, thus

$$\hat{\mu} = \mu^{-1} \mu^{-1} \hat{s}_{ij} = \mu \sum_{ij} \lambda_{ij} \hat{r}_{ij}$$

$$\hat{g}_i = \sum_{ij} g_i \hat{s}_{ij} = \sum_{ij} \hat{g}_i \hat{s}_{ij}$$

Coefficient peculiar to $\mu$.

But since $\hat{s}_{ij} = \sum_{ij} \lambda_{ij} \hat{r}_{ij}$, $\hat{\mu} = \sum_{ij} \lambda_{ij} \hat{r}_{ij}$.

Similarly $\hat{g}_i = \sum_{ij} \lambda_{ij} \hat{s}_{ij}$.

In order to obtain $\hat{s}_{ij}$, use is made of the fact that $\hat{s}_{ij} = \mu + g_i + g_j + s_{ij}$. Therefore, $\hat{s}_{ij} = \hat{s}_{ij} - \mu - g_i - g_j = \sum_{ij} (\lambda_{ij} \hat{r}_{ij} - \lambda_{ij} \hat{r}_{ij}) - \lambda_{ij} \hat{r}_{ij}$. Therefore, $\hat{s}_{ij} = \sum_{ij} \hat{s}_{ij} \hat{r}_{ij}$. 

- 58 -
Similarly \( \hat{r}_{ij} = \hat{r}_{ij} - \hat{\mu} - \hat{e}_i - \hat{e}_j - \hat{m}_j - \hat{s}_{ij} = \sum_{ij} \hat{r}_{ij} \hat{r}_{ij} \).

Now that all of the estimates are expressed as linear functions of the \( \hat{r}_{ij} \) and the variance-covariance matrix of the \( \hat{r}_{ij} \) is available, it is possible to obtain the variance-covariance matrix of the \( \hat{\mu}, \hat{e}_i, \hat{m}_j, \hat{s}_{ij}, \) and \( \hat{r}_{ij} \) from the well known method for finding the variance and covariance of linear functions.

Let 
\[
A = \sum_{i=1}^{p} a_i x_i, \quad i = 1, \ldots, p
\]
\[
B = \sum_{i=1}^{p} b_i x_i
\]
\[
V(x_i) = \sigma_i^2
\]
\[
CV(x_i x_j) = \sigma_{ij}
\]
Then 
\[
V(A) = \sum_{i=1}^{p} a_i^2 \sigma_i^2 + 2 \sum_{i<j=1}^{p} a_i a_j \sigma_{ij}
\]
\[
CV(AB) = \sum_{i=1}^{p} a_i b_i \sigma_i^2 + \sum_{i=1}^{p} a_i b_i \sigma_{ij}
\]

b. \( \hat{e}_i, \hat{m}_j, \hat{s}_{ij}, \hat{r}_{ij} \) are not independent. The procedure for obtaining \( R(\mu, \bar{e}, m, s, a) \) suggests the easiest way for obtaining sampling errors under this assumption. Let \( \hat{s}_{ij} = \mu + \hat{e}_i + \hat{e}_j + \hat{s}_{ij} \) and absorb the coefficients of \( \hat{s}_{ij} \) into the coefficients of \( \hat{m}_j \) and \( \hat{a}_k \). Next invert the reduced matrix and with it compute the entire matrix of least squares coefficients. Then
\[
V(\hat{s}_{ij}) = m_i m_j \sigma_e^2
\]
\[
CV(\hat{m}_i \hat{m}_j) = m_i m_j \sigma_e^2
\]

Since \( \mu \) and \( \bar{e}_i \) are estimated from the \( \hat{s}_{ij} \) by the equations
\[
n_i \mu + \sum n_i \bar{e}_i = \hat{s}_{ii}
\]
\[
n_a \mu + \sum n_a \bar{e}_a = \hat{s}_{a a}
\]
\[\hat{\mu}, \hat{\bar{e}}_i, \) and \( \hat{s}_{ij} \) can be expressed as linear functions of \( \hat{s}_{ij} \). Thus,
\[ \hat{\mu} = \mu^{-1} \mu^{-1} \hat{s} + \sum_i \mu^{-1} g_i \hat{s}_i = \sum_i \lambda_{ij} \hat{s}_{ij}. \]

Similarly, \[ \hat{e}_1 = \sum_{ij} \lambda_{ij} \hat{s}_{ij}, \]

and

\[ \hat{s}_{ij} = \sum_{ij} \lambda_{ij} \hat{s}_{ij}. \]

Since the variances and covariances of the \( \hat{s}_{ij} \) are known, the variances and covariances of \( \hat{\mu}, \hat{e}_1, \) and \( \hat{s}_{ij} \) can be obtained from their expressions as linear functions of \( \hat{s}_{ij} \). Also, the covariances of \( \hat{m}_j \) with \( \hat{\mu}, \hat{e}_1, \) and \( \hat{s}_{ij} \) can be calculated since the covariances among the \( \hat{m}_j \) and the \( \hat{s}_{ij} \) are known.

c. \( E_{ij} E_{ij}^T = m_j = 0 \). Letting \( \tilde{s}_{ij} = \mu + e_1 + e_j + s_{ij} \) and absorbing the coefficients of \( \tilde{s}_{ij} \) into those of \( a_k \), the inverse matrix with respect to \( \tilde{s}_{ij} \) and \( a_k \) is obtained. Then as described in Section II G 3 a and b, the estimates of \( \mu, e_1, \) and \( s_{ij} \) can be expressed as linear functions of the \( \tilde{s}_{ij} \). The variance-covariance matrix of \( \tilde{s}_{ij} \) being known, the variances and covariances of \( \hat{\mu}, \hat{e}_1, \) and \( \hat{s}_{ij} \) are easily obtained.

d. \( E_{ij} E_{ij}^T = m_j = 0 \). As before, \( d_j = \mu + e_j + m_j \). The inverse matrix of \( e_1, d_j, \) and \( a_k \) is obtained after first absorbing the \( d_j \) coefficients. Then \( V(\hat{s}) = e_i e_j^T \sigma_e^2 \), and \( CV(\hat{s}_1, \hat{s}_j) = e_i e_j^T \sigma_e^2 \).

Let \( \tilde{m}_j = \mu + m_j \).

Since \( \tilde{m}_j = d_j - e_j \),

\[ \tilde{m}_j^T \tilde{m}_j = g_i g_j + d_i d_j - 2g_i d_j, \]

\[ \tilde{m}_j^T \tilde{m}_j = g_i e_j + d_i d_j - g_i d_j - g_i d_i \]

\[ g_j^T \tilde{m}_j = g_j d_j - g_i e_j \]

\[ g_i^T \tilde{m}_j = g_i d_j - g_i e_j \]
If the variance-covariance matrix is desired in terms of \( \mu \) and \( m_j \) separately, the transformation described in Section II C 4 b (1) will accomplish that.

\[ \begin{align*}
\hat{z}_1 - \hat{m}_j &= \hat{z}_{ij} - \hat{r}_{ij} = 0.
\end{align*} \]

The inverse matrix involving \( \hat{z}_1 \) and \( \hat{r}_{ij} \) is obtained by first absorbing the \( \mu + a_k \) coefficients. Then

\[ \begin{align*}
V(\hat{z}_1) &= \Gamma_{\hat{z}1}^{-1} \Gamma_{12} \Gamma_{21} \Gamma_{12}^{-1} \Gamma_{22} = \Gamma_{12}^{-1} \Gamma_{22}.
\end{align*} \]

4. Adjustment of the least squares estimates

Since it is assumed that \( \hat{z}_1 \), \( m_j \), \( z_{ij} \), and \( r_{ij} \) are randomly drawn from normal populations, the least squares estimates are not the best estimates of these effects. Rather, the least squares estimates can be thought of as various line or cross-line means measured as deviations from the population mean, each such line or cross line mean having been corrected for the way in which it was associated with other lines and with extraneous circumstances such as degree of inbreeding, age of dam, and season. Now these corrected values are to be used to derive the best possible estimates of the \( \hat{z}_1 \), \( m_j \), \( z_{ij} \), and \( r_{ij} \). What is needed is a regression coefficient for each least squares estimate, say \( b(\hat{z}) = \frac{E(\hat{z})}{E(\hat{z})^2} \).

The needed expected values are easily obtained if the \( \hat{z}_1 \)'s can be expressed as linear functions of the sample observations. This can always be done if necessary, although the computation can be an exceedingly lengthy task.

The procedure will now be applied to the present problem.

\[ \begin{align*}
\hat{z}_1 - \hat{m}_j - \hat{z}_{ij} - \hat{r}_{ij} = 0.
\end{align*} \]

In Section II C 3 a it was shown that \( \hat{z}_1 \), \( \hat{m}_j \), \( \hat{z}_{ij} \), and \( \hat{r}_{ij} \) can be expressed as linear functions of \( \hat{r}_{ij} = \mu + \hat{z}_i + \hat{z}_j + \hat{m}_j + \hat{z}_{ij} + \hat{r}_{ij} \). Thus, \( \hat{z}_a = \sum_{ij} \hat{z}_{a ij} \hat{r}_{ij} \).
In terms of the population parameters \( \hat{r}_{ij} = \mu + \hat{g}_i + \hat{g}_j + m_j + s_{ij} \) 
+ \( r_{ij} + \frac{1}{b} \sum_k a_k + \text{errors}. \)

Therefore, \( \hat{E}_{a} = \frac{p-1}{p} s_a - \frac{1}{p} \sum_{ij} \hat{g}_i + \sum_{ij} (\lambda_{ij} + \lambda_{ji}) s_{ij} + \sum_{ij} \hat{\lambda}_{ij} r_{ij} + \text{errors} \).

\[ \hat{E}_{ab} = \sum_{ij} (s_{ab}^{ij} + s_{ab}^{ji}) s_{ij} + \sum_{ij} \hat{\lambda}_{ij} r_{ij} + \text{errors} \]

\[ \hat{E}_{r_{ab}} = \sum_{ij} \hat{\lambda}_{ij} r_{ij} + \text{errors} \]

The \( b_{\hat{E}_{ab}} \) can now be obtained. They are as follows:

\[ b(\hat{E}_{a}) = \frac{\frac{p-1}{p} \sigma_a^2}{\frac{p-1}{p} \sigma_a^2 + \sum_{ij} (\lambda_{ij} + \lambda_{ji})^2 \sigma_s^2 + \sum_{ij} \hat{\lambda}_{ij} \sigma_r^2 + g_{a}^2 \sigma_e^2} \]

\[ b(\hat{E}_{m}) = \frac{\frac{p-1}{p} \sigma_m^2}{\frac{p-1}{p} \sigma_m^2 + \sum_{ij} \hat{\lambda}_{ij} \sigma_r^2 + m_{a}^2 \sigma_e^2} \]

\[ b(\hat{E}_{s_{ab}}) = \frac{(s_{ab}^{ab} + s_{ab}^{ba}) \sigma_s^2}{\sum_{ij} (s_{ab}^{ij} + s_{ab}^{ji})^2 \sigma_s^2 + \sum_{ij} s_{ab}^{ij} \sigma_r^2 + s_{s_{ab}} \sigma_e^2} \]

\[ b(\hat{E}_{r_{ab}}) = \frac{\hat{\lambda}_{ab}^{ab} \sigma_r^2}{\sum_{ij} \hat{\lambda}_{ij} \sigma_r^2 + r_{ab} \sigma_e^2} \]

b. \( g_{1}, m_{j}, s_{ij} \neq 0; r_{ij} = 0 \). It was shown in Section II 6.3 b that 
\( \hat{g}_i \) and \( \hat{s}_{ij} \) can be expressed as linear functions of \( \hat{r}_{ij} = \mu + \hat{g}_i + \hat{g}_j + \hat{s}_{ij} \).

Thus,
\[ \hat{g}_a = \sum_{ij} \hat{\lambda}_{ij} \hat{s}_{ij} \]
\[ \hat{s}_{ab} = \sum_{ij} \hat{\lambda}_{ij} \hat{s}_{ij} \]
\( \hat{E}_{ij} = \mu + g_i + g_j + \frac{1}{p} \sum_j m_j + s_{ij} + \frac{1}{t} \sum_k a_k + \text{errors. Consequently,} \)

\( \hat{E}_{ia} = \frac{p-1}{p} s_a - \frac{1}{p} \sum_{i \neq i} g_i + \sum_{ij} s_{ij} s_{ij} + \text{errors,} \)

\( \hat{E}_{ij} = \sum_{ij} s_{ij} s_{ij} + \text{errors,} \)

\( \hat{E}_{a} = \frac{p-1}{p} m_a - \frac{1}{p} \sum_{j \neq a} m_j + \text{errors.} \)

Therefore, \( b(s_{a\hat{a}}) = \frac{\frac{p-1}{p} \delta g^2}{\frac{p-1}{p} \delta g^2 + \sum_{ij} s_{aij} s_{aij} + g_{aa} \delta e^2} \).

\( b(m_{a\hat{a}}) = \frac{\frac{p-1}{p} \delta m}{\frac{p-1}{p} \delta m + m_m \delta e^2} \), and

\( b(s_{ab\hat{ab}}) = \frac{\sum_{ij} s_{ab} s_{aij} s_{aij}}{s_{ab} \sum_{ij} s_{aij} s_{aij} + s_{ab} s_{ab} \delta e^2} \).

c. \( g_{i} * s_{ij} \neq 0; m_{j} = r_{ij} = 0. \) \( \hat{g}_{i} \) and \( \hat{s}_{ij} \) can be expressed as linear functions of \( \hat{s}_{ij} = \mu + g_i + g_j + s_{ij} \). Therefore \( b(g_i \hat{g}_{i}) \) and \( b(s_{ij} \hat{s}_{ij}) \) are obtained in the same way as in Section II G 4 b.

d. \( g_{i} * m_{j} \neq 0; s_{ij} = r_{ij} = 0. \) The regressions under this hypothesis can be written without further computation, thus

\( b(g_i \hat{g}_{i}) = \frac{\frac{p-1}{p} \delta g^2}{\frac{p-1}{p} \delta g^2 + g_{i} g_{i} \delta e^2} \).

\( b(m_j \hat{m}_{j}) = \frac{\frac{p-1}{p} \delta m}{\frac{p-1}{p} \delta m + a_{j} a_{j} \delta e^2} \).
e. \( \beta_i \neq 0 \), \( m_j = s_{ij} - r_{ij} = 0 \). This regression can be written at once.

\[
b(\hat{g}_i, \hat{s}_i) = \frac{p-1}{p} \frac{\delta g^2}{\delta g^2 + \delta s^2 + \delta r^2}
\]

5. Estimation of components of variance

Estimates of \( \delta g^2, \delta m^2, \delta s^2, \delta r^2 \), and \( \delta e^2 \) are needed for adjusting the least squares estimates and for planning most efficient programs with respect to making and testing inbred lines. The estimates to be described below are those obtained by taking expected values of the pertinent sums of squares arising in the least squares analysis.

a. Estimation of \( \delta e^2 \). Assuming that \( \delta g^2, \delta m^2, \delta s^2, \delta r^2 > 0 \), the estimate of \( \delta e^2 \) is \( \frac{R(T) - R(\mu, g, m, s, r, a)}{\text{error d.f.}} \). If it should turn out that any of the variances are assumed to be 0, the estimate of \( \delta e^2 \) is \( \frac{R(T) - R(\text{due to parameters} \neq 0)}{\text{error d.f.}} \).

b. Estimation of \( \delta r^2 \). The estimate of \( \delta r^2 \) arises logically from \( R(\mu, g, m, s, r, a) - R(\mu, g, m, s, a) \). The expected value of this sum of squares is \( k_1 \delta r^2 + k_2 \delta e^2 \), where \( k_2 = \text{d.f. for } R \text{ sum of squares and } k_1 = n - \text{the coefficient of } \delta r^2 \text{ in } E R(\mu, g, m, s, a) \). Assuming now that the invert matrix involving \( m_j, \tilde{s}_{ij}, \) and \( s_k \) is available each element of this inverse is multiplied by the coefficient of \( \delta r^2 \) in the expected value of the product of sample sums corresponding to the elements of this matrix.

Once \( k_1 \) is computed, \( \delta r^2 = \frac{1}{k_1} \left[ R(\mu, g, m, s, r, a) - R(\mu, g, m, s, a) - k_2 \delta e^2 \right] \), where \( \delta e^2 \) is the estimate of \( \delta e^2 \) obtained in Section II 5 a.

c. Estimation of \( \delta s^2 \) assuming \( \delta g^2, \delta m^2, \delta r^2 > 0 \). The estimate of \( \delta s^2 \) logically arises from \( R(\mu, g, m, s, a) - R(\mu, g, m, a) \). The expected
value of this sum of squares is 

\[ k_1 \sigma_s^2 + k_2 \sigma_r^2 + k_3 \sigma_e^2, \]

where

\[ k_1 = n - \text{coefficient of } \sigma_s^2 \text{ in } E R(\mu, g, m, a), \]

\[ k_2 = \text{coefficient of } \sigma_r^2 \text{ in } E R(\mu, g, m, s, a) - \text{coefficient of } \sigma_r^2 \text{ in } E R(\mu, g, m, a), \]

\[ k_3 = \text{degrees of freedom for S sum of squares}. \]

Then

\[ \hat{\sigma}_s^2 = \frac{1}{k_1} \left[ R(\mu, g, m, s, a) - k_2 \hat{\sigma}_r^2 - k_3 \hat{\sigma}_e^2 \right], \]

\[ \hat{\sigma}_r^2 \text{ and } \hat{\sigma}_e^2 \text{ being the estimates obtained in Section II G 5 a and Section II G 5 b respectively.} \]

d. Estimation of \( \sigma_s^2 \) assuming \( \sigma_r^2 > 0; \sigma_e^2 = 0 \). The procedure is now simplified as compared to Section II G 5 c for \( E R(\mu, g, m, s, a) - R(\mu, g, m, a) = k_1 \sigma_s^2 + k_3 \sigma_e^2 \). \( k_1 \) and \( k_3 \) have the same values as in Section II G 5 c.

Now

\[ \hat{\sigma}_s^2 = \frac{1}{k_1} \left[ R(\mu, g, m, s, a) - R(\mu, g, m, a) - k_2 \hat{\sigma}_e^2 \right], \]

where

\( \hat{\sigma}_e^2 \) is the error d.f. for

\[ \frac{1}{k_1} \left[ \frac{R(T) - R(\mu, g, m, s, a)}{} \right]. \]

e. Estimation of \( \sigma_s^2 \) assuming \( \sigma_e^2 > 0; \sigma_r^2 = 0 \). The estimate of \( \sigma_s^2 \) now arises from \( R(\mu, g, s, a) - R(\mu, g, a) \), the expected value of which is \( k_1 \sigma_s^2 + k_2 \sigma_e^2 \), where \( k_1 = n \) minus the coefficient of \( \sigma_s^2 \)

in \( R(\mu, g, a) \) and \( k_2 = \text{degrees of freedom for S sum of squares}. \)

Then

\[ \hat{\sigma}_s^2 = \frac{1}{k_1} \left[ R(\mu, g, s, a) - R(\mu, g, a) - k_2 \hat{\sigma}_e^2 \right], \]

where

\[ \hat{\sigma}_e^2 = \frac{1}{k_1} \left[ R(T) - R(\mu, g, s, a) \right]. \]

f. Estimation of \( \sigma_m^2 \) assuming \( \sigma_r^2, \sigma_s^2, \sigma_e^2 > 0 \). This estimate comes from \( R(\mu, g, m, s, a) - R(\mu, g, s, a) \), the expected value of which is \( k_1 \sigma_m^2 + k_2 \sigma_r^2 + k_3 \sigma_e^2 \), where

\[ k_1 = n - \text{coefficient of } \sigma_m^2 \text{ in } E R(\mu, g, s, a), \]

\[ k_2 = \text{coefficient of } \sigma_r^2 \text{ in } E R(\mu, g, m, s, a) - \text{coefficient of } \sigma_r^2 \text{ in } E R(\mu, g, m, a). \]
\[ \sigma_r^2 \text{ in } E R(\mu, g, s, a) \]

\[ k_3 = \text{degrees of freedom for } M \text{ sum of squares} = p-1 \]

Then \[ \hat{\sigma}_m^2 =\frac{1}{k_1} \left[ R(\mu, g, m, s, a) - R(\mu, g, s, a) - k_2 \hat{\sigma}_r^2 - k_3 \hat{\sigma}_e^2 \right], \]

where \( \hat{\sigma}_r^2 \) is the same as in Section II G 5 b and \( \hat{\sigma}_e^2 \) the same as in Section II G 5 a.

**g. Estimation of \( \sigma_m^2 \) assuming \( \sigma_g^2, \sigma_s^2 > 0; \sigma_r^2 = 0 \).** Again \( R(\mu, g, m, s, a) - R(\mu, g, s, a) \) is used. The expected value of this sum of squares is \( k_1 \sigma_m^2 + k_3 \sigma_e^2 \), and \( \hat{\sigma}_m^2 = \frac{1}{k_1} \left[ R(\mu, g, m, s, a) - R(\mu, g, s, a) - k_3 \hat{\sigma}_e^2 \right] \). \( k_1 \) and \( k_3 \) are the same as in Section II G 5 f, but \( \hat{\sigma}_e^2 \) is the estimate coming from \( R(T) - R(\mu, g, m, s, a) \).

**h. Estimation of \( \sigma_m^2 \) assuming \( \sigma_g^2 > 0; \sigma_s^2 = \sigma_r^2 = 0 \).** The sum of squares used for this estimate is \( R(\mu, g, m, a) - R(\mu, g, a) \), the expected value of which is \( k_1 \sigma_m^2 + k_2 \sigma_e^2 \).

\[ k_1 = n - \text{coefficient of } \sigma_m^2 \text{ in } E R(\mu, g, m) \]

\[ k_2 = p - 1. \]

\[ \hat{\sigma}_m^2 = \frac{1}{k_1} \left[ R(\mu, g, m, a) - R(\mu, g, a) - k_2 \hat{\sigma}_e^2 \right], \]

where \( \hat{\sigma}_e^2 = \frac{1}{\text{error d.f.}} \left[ R(T) - R(\mu, g, m, a) \right]. \)

**i. Estimation of \( \sigma_g^2 \) assuming \( \sigma_m^2, \sigma_s^2, \sigma_r^2 > 0 \).** For this estimate the sums of squares, \( R(\mu, g, m, a) - R(\mu, m, a) \), whose expected value is \( k_1 \sigma_g^2 + k_2 \sigma_s^2 + k_3 \sigma_r^2 + k_4 \sigma_e^2 \) is utilized.

\[ k_1 = 2n - \text{coefficient of } \sigma_g^2 \text{ in } E R(\mu, m, a) \]

\[ k_2 = \text{coefficient of } \sigma_s^2 \text{ in } E R(\mu, g, m, a) - \text{coefficient of } \sigma_g^2 \text{ in } E R(\mu, m, a) \]

\[ k_3 = \text{coefficient of } \sigma_r^2 \text{ in } E R(\mu, g, m, a) - \text{coefficient of } \sigma_r^2 \text{ in } E R(\mu, m, a) \]

\[ k_4 = p - 1 \]
Then \( \hat{\sigma}_g^2 = \frac{1}{k_1} \left[ R(\mu, g, m, a) - R(\mu, g, a) - k_2 \hat{\sigma}_s^2 - k_3 \hat{\sigma}_r^2 - k_4 \hat{\upsilon}_e^2 \right] \),
where \( \hat{\sigma}_s^2 \), \( \hat{\sigma}_r^2 \), and \( \hat{\upsilon}_e^2 \) are the estimates already made under the assumption that all variances are greater than zero.

j. Estimation of \( \hat{\sigma}_g^2 \) assuming \( \hat{\sigma}_m^2, \hat{\sigma}_s^2 > 0 \); \( \hat{\sigma}_r^2 = 0 \). The estimate of \( \hat{\sigma}_g^2 \) in this case is made similarly to that in Section II G 5 i except that \( \hat{\sigma}_r^2 \) is deleted, and the estimates of \( \hat{\sigma}_s^2 \) and \( \hat{\upsilon}_e^2 \) utilized are the ones arising from the assumption that \( \hat{\sigma}_s^2 > 0; \hat{\sigma}_r^2 = 0 \).

k. Estimation of \( \hat{\sigma}_g^2 \) assuming \( \hat{\sigma}_s^2 > 0; \hat{\sigma}_m^2 = \hat{\sigma}_r^2 = 0 \). The estimate of \( \hat{\sigma}_g^2 \) under this assumption is made the same as was the estimate in Section II G 5 i except that \( \hat{\sigma}_s^2 \) and \( \hat{\sigma}_r^2 \) are dropped from the equations, and
\[
\hat{\upsilon}_e^2 = \frac{1}{\text{error d.f.}} \left[ R(T) - R(\mu, g, m, a) \right].
\]

l. Estimation of \( \hat{\sigma}_g^2 \) assuming \( \hat{\sigma}_s^2 > 0; \hat{\sigma}_m^2 = \hat{\sigma}_r^2 = 0 \). This estimate arises from \( R(\mu, g, a) - R(\mu, a) \), the expected value of which is \( k_1 \hat{\sigma}_g^2 + k_2 \hat{\sigma}_s^2 + k_3 \hat{\upsilon}_e^2 \), where
\[
k_1 = 2n - \text{coefficient of } \hat{\sigma}_g^2 \text{ in } \mathbb{E} R(\mu, a)
\]
\[
k_2 = \text{coefficient of } \hat{\sigma}_s^2 \text{ in } \mathbb{E} R(\mu, g, a) - \text{coefficient of } \hat{\sigma}_s^2 \text{ in } \mathbb{E} R(\mu, a)
\]
\[
k_3 = p - 1.
\]
\[
\hat{\sigma}_g^2 = \frac{1}{k_1} \left[ R(\mu, g, a) - R(\mu, a) - k_2 \hat{\sigma}_s^2 - k_3 \hat{\upsilon}_e^2 \right], \text{ where } \hat{\sigma}_s^2 \text{ and } \hat{\upsilon}_e^2
\]
are the estimates obtained under the assumption that \( \hat{\sigma}_s^2 > 0; \hat{\sigma}_m^2 = \hat{\sigma}_r^2 = 0 \).

m. Estimation of \( \hat{\sigma}_g^2 \) assuming \( \hat{\sigma}_s^2 = \hat{\sigma}_m^2 = \hat{\sigma}_r^2 = 0 \). This estimate is effected in the same manner as was the estimate of \( \hat{\sigma}_g^2 \) in Section II G 5 i except that \( \hat{\sigma}_s^2 \) is deleted and
\[
\hat{\upsilon}_e^2 = \frac{1}{\text{error d.f.}} \left[ R(T) - R(\mu, g, a) \right].
\]
6. Prediction of line cross performance

Given that a least squares analysis has been made of the data arising from single crosses among p inbred lines, what use can be made of such estimates in predicting the outcome of future crosses? The following questions can be answered:

1. What is the most probable value of progeny of males of a tested inbred line when mated to a random sample of females from the same population of lines?

2. What is the most probable value of progeny of females of the different tested inbred lines when mated to a random sample of males from the same population of lines?

3. What is the most probable value of progeny of a specific cross which has been tested reciprocally?

4. What is the most probable value of progeny of a specific cross which has been tested in only one of the two possible reciprocal crosses?

5. What is the most probable value of progeny of a specific cross between two lines which have not previously been tested in crosses with each other?

The problem is the one discussed in Section II E of estimating a parameter which is itself a linear function of certain randomly drawn effects, given least squares estimates of these effects. The parameter to be estimated in this case is the most probable value of a particular cross. Let this parameter be $T_{ij}$, the $i$ and $j$ denoting progeny of the $i^{th}$ line of sire mated to the $j^{th}$ line of dam. In terms of parameters, $T_{ij} = g_i + g_j$. 
What is needed is an index $I_{ij}$ of the form, $I_{ij} = b_1 \hat{g}_1 + b_2 \hat{g}_2 + b_3 \hat{m}_j + b_4 \hat{g}_1 \hat{m}_j + b_5 \hat{r}_{ij}$, to be used in predicting from least squares estimates the most probable value of a specific cross.

a. Performance as line of sire. The probable average performance of the progeny of an inbred line used as the male line is $g_1$. The appropriate index is not, however, $b \hat{g}_1$ as might be expected. Instead it is $b_1 \hat{g}_1 + b_2 \hat{m}_1$. The reason for this is that $\hat{g}_1$ and $\hat{m}_1$ are correlated, and consequently $\hat{m}_1$ tells something about $g_1$. This correlation comes about since $\hat{m}_1 = \hat{d}_1 - \hat{\mu} - \hat{\hat{g}_1}$. An additional factor which would cause correlation between $\hat{g}_1$ and $\hat{m}_1$ would be correlation between $g_1$ and $m_1$. The solution to the following set of equations properly takes into account both of these causes of correlation.

\[
\begin{bmatrix}
  b_1 & b_2 & 1 \\
  \hat{E}g_1^2 & \hat{E}g_1 \hat{m}_1 & \hat{E}g_1 \\
  \hat{E}g_1 \hat{m}_1 & \hat{E}m_1^2 & \hat{E}m_1 \\
\end{bmatrix}
\]

The expected values needed in the preceding set of equations and in succeeding ones are computed from the expected values of the least squares estimates (Section II G 4) and the linear function describing $T$, the true value, remembering that $\hat{E}g_1 \hat{m}_1 = \hat{\sigma}_{gm}$.

It should be pointed out that the assumption that $\hat{E}(g_1 m_1) \neq 0$ does not affect the estimates of the variance components already described. This is true because the coefficients of $\hat{\sigma}_{gm}$ in the expectation of any reduction in which either $g_1$ or $m_j$ or both $g_1$ and $m_j$ are fitted equals twice the total
number of observations. But all of the reductions utilized in estimating
components of variance have in them either or both \( g_i \) and \( m_j \), and, consequently, \( \sigma_{gm} \) goes out in the difference between the two reductions.

Many ways could be devised for obtaining unbiased estimates of \( \sigma_{gm} \).
A very simple one involves calculating \( \sum_i \hat{g}_i m_i \). The expected value of
this quantity is
\[
\sum_i E(\hat{g}_i m_i) = \sum_i \left( \frac{p-1}{p} \sigma_{gm} + \sum_{ij} \lambda_{ij} m_i \lambda_{ij} \sigma_r^2 + g_i^m \sigma_e^2 \right)
\]
\[
= (p-1) \sigma_{gm} + \sum_{ij} \left[ \frac{\sigma_r^2 + g_i^m \sigma_e^2}{m_i} \right].
\]

Therefore, if \( \sum_i \hat{g}_i m_i \) is first corrected for the \( \sigma_r^2 \) and \( \sigma_e^2 \) com-
ponents and then divided by \( p-1 \), an unbiased estimate of \( \sigma_{gm} \) is obtained.
A better estimate might be obtained by weighting each of the \( \hat{g}_i m_i \) by some
factor related to the confidence which is placed in the estimates of \( g_i \)
and \( m_i \).

b. Performance as line of dam. The probable average performance as
line of dam is \( T_{ij} = g_j + m_j \), and the index needed is \( I = b_1 \hat{g}_j + b_2 \hat{m}_j \).
The equations for computing \( b_1 \) and \( b_2 \) are

\[
\begin{array}{ccc}
b_1 & b_2 & 1 \\
E \hat{g}_j^2 & E \hat{g}_j \hat{m}_j & E g_j \hat{g}_j + E \hat{g}_j m_j \\
E \hat{g}_j \hat{m}_j & E \hat{m}_j^2 & E g_j \hat{m}_j + E m_j \hat{m}_j \\
\end{array}
\]

e. Performance of specific crosses previously tested reciprocally.
The expected performance of a particular cross is \( T_{ij} = g_i + g_j + m_j +
\( s_{ij} + r_{ij} \), and the index needed is \( I = b_1 \hat{g}_1 + b_2 \hat{g}_j + b_3 \hat{m}_j +
b_4 \hat{s}_{ij} + b_5 \hat{r}_{ij} \).
The necessary equations are
The lower left hand corner of the matrix of coefficients is left blank because it is symmetric with the upper right hand corner.

It is quite apparent that \( I = b_1 \hat{s}_1 + b_2 \hat{e}_j + b_3 \hat{m}_j + b_4 \hat{s}_{ij} + b_5 \hat{r}_{ij} \) yields a better prediction of the results of a cross of the \( i^{th} \) line of sire by the \( j^{th} \) line of dam than does the estimate arising from \( \bar{y}_{ij} \), as the latter utilizes only \( n_{ij} \) sample observations, while I utilizes, in addition, all the data in which either the \( i^{th} \) or the \( j^{th} \) line appears. The index, \( I \), properly weights the different sources of information according to the number of observations and according to the relative sizes of the different variances. The difficulty is that in other than balanced experiments the computation of the index for each of the possible crosses is a formidable task indeed.

d. **Performance of specific crosses previously tested in only one way.**

Since there is now available no estimate of \( r_{ij} \), it is assumed to be zero. The equations are therefore the same as in Section II G 6 a except that the last column and row are deleted.

e. **Performance of specific crosses not previously tested.** There is
now available no estimate of either $s_{ij}$ or $r_{ij}$, and they are both, therefore, assumed to be zero. The necessary equations for constructing the index are the same as in Section II 6 c except that the last two rows and last two columns are deleted.

Any of the above indexes can be modified by making different assumptions than those made here. For example, if it is assumed that $r_{ij} = \sigma_r^2 = 0$, everything pertaining to $r_{ij}$ and $\sigma_r^2$ drops out of the equations and out of the expected values used in setting up the coefficients and the right members of the equations. If it is assumed that $\sigma_{g_m} = 0$, this covariance drops out of all of the expected values.

H. Extension of the Methods to More Complex Tests of Lines

1. Consideration of differences among sires

In other than nearly homozygous lines, sire differences within lines may be appreciable. Therefore, apparent line differences may be due in part to accidents of sampling consequent to the selection of the sire or sires used in the test of a line. If more than one sire is used per line per year in the single cross tests, estimates can be obtained of sire differences and sire by line of dam interaction. The mathematical model is now modified as follows,

$$y_{ijklm} = \mu + g_i + s_j + m_k + s_{ij} + r_{ij} + b_{ik} + b_{ijk} + a_1 + e_{ijklm}$$

$b_{ik}$ is an effect common to progeny of the $k^{th}$ male of the $i^{th}$ line.

$b_{ijk}$ is an effect common to progeny of crosses of the $k^{th}$ male of the $i^{th}$ line by females of the $j^{th}$ line.

Tests of significance and estimation of variances prove very difficult
to accomplish because of the large number of interaction terms. For example, the formal test of the hypothesis that $b_{ijk} = 0$ involves calculation of $R(\mu, g, m, s, r, b, b_1, a) - R(\mu, g, m, s, r, b, a)$. The first of these reductions is easily obtained by absorbing boar $x$ line of dam subclass constants. The second, however, could require many equations for it seems that the shortest method involves absorbing the reciprocal cross coefficients, thereby leaving a set of equations in $b_{ik}$ and $a_1$.

The test of the hypothesis that $b_{ik} = 0$ assuming that $b_{ijk} = 0$, involves $R(\mu, g, m, s, r, b, a) - R(\mu, g, m, s, r, a)$. Both of these reductions have already been described, the first in the paragraph above and the second in Section II G 1 b. If $r_{ij}$ as well as $b_{ijk}$ is assumed $= 0$, the test is $R(\mu, g, m, s, b, a) - R(\mu, g, m, s, a)$. It appears that the best way to obtain the first of these two reductions is to absorb $\mu + g_i + g_j + s_{ij}$. It should be noted that the portion of the reduced equations involving the $b_{ik}$ should be set up so as to retain the symmetry of the coefficients. The test of the hypothesis that $b_{ik} = 0$ is most easily accomplished, however, if the assumption is made that $r_{ij} = s_{ij} = b_{ijk} = 0$.

Then the sum of squares for $B = R(\mu, g, m, b, a) - R(\mu, g, m, a)$. The first reduction is best computed by letting $b_{ik} = \mu + g_i + b_{ik}$ and $d_j = g_j + m_j$. Then the $b_{ik}$ coefficients can be absorbed. The second reduction was described in Section II G 1 b. Under the assumption that $m_j = s_{ij} = r_{ij} = b_{ijk} = 0$, the test is quite difficult. The $B$ sums of squares = $R(\mu, g, b, a) - R(\mu, g, a)$. It does not appear possible to absorb either the $g_i$ or $b_{ik}$. It is again necessary when reducing these equations and unknowns to retain the symmetry of the $b_{ik}$ portion of the matrix.
2. Consideration of differences among dams

In the usual situation of line testing in swine, each sow would raise only one litter, and consequently the analysis would need to be on an intra-line of dam intra-sire basis. The model previously given would be modified as follows,

\[ y_{ijklmn} = \mu + e_i + e_j + a_{ij} + r_{ij} + b_{ik} + b_{ijk} + d_{ijkl} + a_m + e_{ijklmn} \]

\( d_{ijkl} \) is an effect common to progeny of the \( i^{th} \) dam of the \( j^{th} \) line mated to the \( k^{th} \) male of the \( l^{th} \) line.

The sum of squares for dams = \( R(\mu, e_i, e_j, a_{ij}, r_{ij}, b_{ik}, b_{ijk}, d_{ijkl}, a_m) \) - \( R(\mu, e_i, e_j, a_{ij}, r_{ij}, b_{ik}, b_{ijk}, d_{ijkl}) \). The first reduction is most easily obtained by letting \( \hat{d}_{ijkl} = \mu + e_i + e_j + a_{ij} + r_{ij} + b_{ik} + b_{ijk} + d_{ijkl} \). The model then reduces to \( y_{ijklmn} = \hat{d}_{ijkl} + a_m + e_{ijklmn} \), and the least squares equations are,

\[ n_{ijkl} \hat{d}_{ijkl} + \sum_m n_{ijkl} a_m = y_{ijkl} \]
\[ \sum_{ijkl} n_{ijkl} \hat{d}_{ijkl} + n \sum_m a_m = y \]

The coefficients of \( \hat{d}_{ijkl} \) can be absorbed into those of \( a_m \). Computation of the second reduction was described in Section II H 1.

In certain species it might be possible to obtain two or more progeny from a single dam out of each of two or more sires. Then it is feasible to obtain estimates of dam by line of sire interaction and of sire by dam interaction.

3. Breed crosses

The occasion might arise for testing inbred lines from several different
breeds in single crosses, the crosses being made only among breeds. For example, several stations might develop lines of only one or two breeds, conduct intra-breed cross line tests, and then submit selected lines for tests of inter-breed crosses at some central testing station. A mathematical model applicable to this design is,

\[ y_{ijklm} = \mu + a_i + b_j + c_{ij} + g_{ik} + e_{jl} + m_{jl} + s_{ijkl} + e_{ijklm} \]

where \( y_{ijklm} \) is the \( m \)-th unit of matings between males of the \( k \)-th line of the \( i \)-th breed by females of the \( l \)-th line of the \( j \)-th breed, \( a_i \) (\( a_j \)) is an effect common to progeny of the \( i \)-th (\( j \)-th) breed, \( b_j \) is an effect common to progeny of the \( j \)-th breed when this breed is used as line of dam, \( c_{ij} \) is an effect common to progeny of crosses between the \( i \)-th and \( j \)-th breeds, \( g_{ik} \) (\( g_{jl} \)) is an effect common to progeny of the \( k \)-th line of the \( i \)-th breed (\( l \)-th line of the \( j \)-th breed), \( m_{jl} \) is an effect common to progeny of females of the \( l \)-th line of the \( j \)-th breed, and \( s_{ijkl} \) is an effect common to the progeny of crosses between the \( k \)-th line of the \( i \)-th breed by the \( l \)-th line of the \( j \)-th breed.

From such data mean squares can be calculated for the followings:

1. General combining ability among breeds.
2. Maternal ability among breeds.
3. Specific combining ability among breeds.
4. General combining ability among lines within breeds.
5. Maternal ability among lines within breeds.
6. Specific combining ability of line crosses within breeds.

4. Three-way crosses

One of the possible ways for utilizing inbreds commercially is the
cross of inbred males on single cross females. Experimental data from such crosses can furnish information concerning the general combining ability of lines, the general combining ability for maternal ability in single cross females, specific combining ability with respect to performance as single cross dams, and three-way interaction or specific combining ability. A mathematical model from which such information can be obtained is

\[ y_{ijkl} = \mu + \frac{1}{2} g_i + \frac{1}{2} g_j + \frac{1}{2} g_k + \frac{1}{2} m_j + \frac{1}{2} m_k + s_{jk} + t_{ijk} + e_{ijkl} \]

\( y_{ijkl} \) is the measurement on the \( i \)th unit in matings of males of the \( i \)th line with \( j \times k \) (or \( k \times j \)) single cross females. \( g_i \) is comparable to the \( g_i \) in the single cross model, \( m_j \) is comparable to the \( m_j \) in the single cross model, \( s_{jk} \) is an effect peculiar to progeny of single cross females of \( j \)th \( x \) \( k \)th (or \( k \)th \( x \) \( j \)th) lines, and \( t_{ijk} \) is an effect peculiar to the progeny of \( i \)th line of males by \( j \) \( x \) \( k \) (or \( k \) \( x \) \( j \)) single cross females.

The least squares equations for this model are as follows:

\[ \hat{\mu} = \frac{\sum (n_{i..} + \frac{1}{2} n_{..i} \cdot g_i + \frac{1}{2} \sum j n_{..j} \cdot m_j + \sum j k n_{..jk} \cdot s_{jk} + \sum i j k t_{ijk} = y_{...}}{n...} \]

\[ g_a = \left( n_{a..} + \frac{1}{2} n_{..a} \right) \hat{\mu} + \left( n_{a..} + \frac{1}{2} n_{..a} \right) g_a + \sum j n_{a..} \cdot m_j + \sum j k n_{a..} \cdot s_{jk} + \sum i j k n_{a..} \cdot t_{ijk} + y_{a..} \]

\[ m_a = \left( \frac{1}{2} n_{a..} \right) \hat{\mu} + \left( \frac{1}{2} n_{a..} \right) g_a + \sum i n_{a..} \cdot g_i + \sum j m_{a..} \cdot m_j + \sum j k m_{a..} \cdot s_{jk} + \sum i j k m_{a..} \cdot t_{ijk} = y_{a..} \]
\[ s_{ab} = n_{ab} (\mu + \frac{1}{b} \varepsilon_a + \frac{1}{a} \varepsilon_b + \frac{1}{c} \mu_a + \frac{1}{d} \mu_b + \varepsilon_{ab}) + \sum_i n_{iab} \xi_i \]
\[ + \sum_i n_{iab} t_{iab} = y_{ab}. \]
\[ t_{abc} = n_{abc} (\mu + \varepsilon_a + \frac{1}{b} \varepsilon_b + \frac{1}{c} \varepsilon_c + \frac{1}{d} \mu_a + \frac{1}{e} \mu_b + \mu_c + s_{bc} + t_{abc}) = y_{abc}. \]

In solving the least squares equations the usual restrictions are imposed, thus,
\[ \sum_i \xi_i = \sum_j m_j = \sum_k s_{jk} \quad \text{(for all \( j \))} \]
\[ = \sum_i t_{ijk} \quad \text{(for all \( i, k \))} \]
\[ = \sum_j t_{ijk} \quad \text{(for all \( j \))} = 0. \]

a. Test of hypothesis that \( t_{ijk} = 0 \). A logical starting place in the analysis of the three way cross data is to test the hypothesis that \( t_{ijk} = 0 \). The sum of squares for \( R = R(\mu, \varepsilon, m, s, t) - R(\mu, \varepsilon, m, s) \). The first of these reductions can be very easily obtained if it is noted that in the \( t_{ijk} \) equations all of the parameters can be combined into a joint parameter say, \( \tilde{t}_{ijk} = \mu + \varepsilon_i + \frac{1}{b} \varepsilon_j + \frac{1}{c} \varepsilon_k + \frac{1}{d} \mu_j + \frac{1}{e} \mu_k + s_{jk} + t_{ijk} \). The number of independent parameters is exactly equal to the number of \( t_{ijk} \) equations. Consequently, the only equations to be solved are \( n_{ijk} \tilde{t}_{ijk} = y_{ijk} \). Consequently, \( R(\mu, \varepsilon, m, s, t) = \sum_{ijk} \frac{y_{ijk}^2}{n_{ijk}} \).

Computation of \( R(\mu, \varepsilon, m, s) \) is facilitated if the joint parameter \( \tilde{e}_{jk} = \mu + \frac{1}{b} \mu_j + \frac{1}{c} \mu_k + s_{jk} \) is utilized. Then the \( g_j \) and \( \tilde{e}_{jk} \) equations number one more than the number of independent parameters. They are reduced to the correct number by utilizing \( \sum_i g_i = 0 \). The equations to be solved are, therefore,
The coefficients of the $e_{ab}$ can be absorbed into the $e_1$ coefficients, thereby reducing the equations to be solved to $p-1$. Then $R(\mu, g, m, s) = \sum_i \hat{e}_i (y_{i...} + \frac{1}{2} y_{i..}) + \sum_{jk} \hat{e}_{jk} y_{..jk}$. The degrees of freedom for the $T$ sum of squares $= u - v - p + 1$, where $u$ = number of different three-way crosses, $v$ = number of different single crosses used as females, and $p$ = number of lines used as the male line.

The error sum of squares is $R(\text{Total}) - R(\mu, g, m, s, t)$, with $n... - u$ degrees of freedom.

All subsequent tests will be described as though the hypothesis that $t_{ijk} = 0$ is accepted.

b. Test of hypothesis that $e_{jk} = 0$. The test of the hypothesis that $e_{jk} = 0$ appears to be the next logical step in the analysis. The sum of squares for $S$ is $R(\mu, g, m, s) - R(\mu, g, m)$. The first of these reductions has already been described. The second is most easily obtained by estimating the joint parameters $d_j = e_j + m_j$, and $\hat{e}_i = u + e_i$. The equations then reduce to,

$$\hat{e}_a \quad n_a \cdot \hat{e}_a + \frac{1}{2} \sum_j n_{aj} \cdot d_j = y_{a...}$$

$$d_a \quad \frac{1}{2} \sum_i n_{ia} \cdot \hat{e}_i + \frac{1}{2} n_a \cdot d_a + \sum_j \frac{1}{2} n_{aj} \cdot d_j = \frac{1}{2} y_{a..}$$

The coefficients of the $\hat{e}_a$ can now be absorbed into those of the $d_j$, thereby reducing the equations to ones involving only $p-1$ unknowns. Then
The degrees of freedom for $S$ sum of squares $= v - p$.

The error sum of squares is $R(\text{Total}) - R(\mu, g, m, a)$ with $n - v - p + 1$ degrees of freedom if it is assumed that $t_{ijk} = 0$. If $t_{ijk} \neq 0$, the error degrees of freedom are the same as for the test of the hypothesis that $t_{ijk} = 0$. In the latter case the $S$ sum of squares has some bias in it, but the removal of it by determining $R(\mu, g, m, s, t) - R(\mu, g, m, t)$ is extremely laborious.

c. Test of hypothesis that $m_j = 0$. In order to make this test with any reasonable degree of facility, it is necessary to assume that $t_{ijk} = s_{jk} = 0$. The $M$ sum of squares is then $R(\mu, g, m) - R(\mu, g)$. The first of these reductions was described in Section II 4 b. The second is obtained by solving the following equations:

$$
\hat{e}_a = n_a \cdot \frac{1}{2} n_a + \frac{1}{2} \sum_j (\frac{1}{2} n_{aj} + \frac{1}{2} n_{ja} + \frac{1}{2} n_{aj}) \hat{e}_j = y_a . . .
$$

Then

$$
R(\mu, g) = \sum_i \hat{e}^2_i (y_i . . . + \frac{1}{2} y_i . . .) .
$$

Assuming that $t_{ijk} = s_{jk} = 0$, the error sum of squares is $R(\text{Total}) - R(\mu, g, m)$ with $n - 2p + 1$ degrees of freedom. The test could also be made assuming $g_1 = 0$, in which case the sum of squares for $M = R(\mu, m) - R(\mu)$.

d. Test of hypothesis that $g_1 = 0$. This test can be made under any of the following assumptions.

1. $m_j, s_{jk} \neq 0$ in which case the sums of squares for $G = R(\mu, g, m, s) - R(\mu, m, s)$.

2. $m_j \neq 0; s_{jk} = 0$. The sum of squares in this case is $R(\mu, g, m) - R(\mu, m)$. 


3. \( m_j = s_{jk} = 0 \). The sum of squares for G in this simplified case = \( R(\mu, g) - R(\mu) \).

\( R(\mu, m, s) \) can be obtained easily by solving the equations, \( n_{jk} \tilde{s}_{jk} = y_{jk} \), where \( \tilde{s}_{jk} = \mu + \frac{1}{k} m_j + \frac{1}{k} m_k + s_{jk} \). Then \( R(\mu, m, s) = \sum_{jk} \frac{1}{n_{jk}} y^2_{jk} \).

\( R(\mu, g, m, s) \) and \( R(\mu, g, m) \) were described in Section II H 4 a and II H 4 b respectively.

\( R(\mu, m) \) can be obtained by solving the equations, \( \frac{1}{n_{a}} \tilde{m}_a + \sum_j \frac{1}{n_{aj}} n_{aj} \).

\( \tilde{m}_j = \frac{1}{k} y_{a...} \), where \( \tilde{m}_j = \mu + m_j \).

Then \( R(\mu, m) = \frac{1}{n} \sum_j \tilde{m}_j y_{a...} \).

\( R(\mu) = \frac{\sum y^2_{a...}}{n} \).

The calculation of parameter estimates, sampling errors, variance estimates, and regression of least squares estimates involves no principles that have not been discussed in connection with single crosses. The same applies to other criteria of classification which might be added such as age and degree of inbreeding.

I. Some Balanced Designs for Testing Lines

Completely balanced designs for testing inbred lines in various kinds of tests have been developed in this study. The designs include single crosses, top crosses, farm tests, three-way crosses, and own performance. Computational procedures for estimating parameters, for obtaining sampling errors of the estimates, and for carrying out tests of hypotheses have been
developed. In addition, computational procedures for estimating variance components and for adjusting the least squares estimates have been obtained for the single cross design.

Such balanced designs are needed for two different purposes.

1. For comparing the efficiency of different types of tests.
2. For analyzing data obtained from a balanced experiment.

The computational procedures to be described illustrate very vividly the great saving in labor which can be effected if the designs are completely balanced. Also, troublesome questions such as what assumptions to make regarding the presence or absence of interaction have no bearing on the estimates of general combining ability or maternal effects. Furthermore, the information obtainable on each of the lines is of equal precision.

It would seem possible in the case of some species to set up completely balanced designs in order to utilize the easier computational procedure and the other advantages of such designs. If the design should lack only a small amount of being completely balanced, it might well be feasible to discard enough of the data to balance the design. The saving in time thereby effected could conceivably more than compensate for the loss of information arising from failure to utilize all of the data, particularly if decisions with respect to breeding plans awaited the results of the analysis.

1. Single crosses

Assume that there are p inbred lines. A completely balanced design for testing these lines in single crosses is one in which data are available on n individuals (litters) of each of the p(p-1) possible reciprocal crosses. Furthermore, each of the reciprocal crosses is orthogonal to any measurable
Confounding factors such as coefficient of inbreeding and season, or the
data are corrected for such differences on the basis of a priori informa-
tion.

The mathematical model used is the one previously described. That is,
\[ y_{ijk} = \mu + \xi_i + \varepsilon_j + m_j + a_{ij} + r_{ij} + e_{ijk} \]

a. Least squares estimates. The least squares equations are
\[
\begin{align*}
\mu: & \quad p(p-1) n \mu + 2(p-1) n \sum g_i + (p-1) n \sum m_j + 2n \sum s_{ij} + n \sum r_{ij} \\
\varepsilon_j: & \quad 2(p-1) n \mu + 2(p-1) n \varepsilon_j + 2n \sum g_i + (p-1) n m_a + n \sum m_j \\
& \quad + 2n \sum s_{aj} + n \sum r_{ia} + n \sum r_{aj} = y_{aj} + y_{a..} \\
m_a: & \quad (p-1) n \mu + (p-1) n \varepsilon_j + n \sum g_i + (p-1) n m_a + n \sum s_{ia} \\
& \quad + n \sum r_{ia} = y_{ia} \\
s_{ab}: & \quad 2n (\mu + g_a + g_b) + n (m_a + m_b) + 2n s_{ab} + n (r_{ab} + r_{ba}) = y_{ab} + y_{ba} \\
r_{ab}: & \quad n (\mu + g_a + g_b + m_a + m_b + s_{ab} + r_{ab}) = y_{ab}.
\end{align*}
\]

Considering now the matrix of coefficients in the complete set of
least squares equations it can be seen that if use is made of the follow-
ing restrictions, \( \sum g_i = \sum m_j = \sum s_{aj} = \sum r_{aj} = \sum r_{ia} = 0 \), all parameters
save \( \mu \) drop out of the \( \mu \) equations, the \( s_{ij} \) and \( r_{ij} \) drop out of the \( g_i \) and
\( m_j \) equations, and the \( r_{ij} \) drop out of the \( s_{ij} \) equations. Then if certain
equations are subtracted from others in such a way as to maintain the
symmetry of the original equations, the matrix of coefficients of the inde-
dendent parameters has in it the following types of terms:
\[ \mu \mu: \ p(p-1)n \quad m_i m_j: \ 2(p-1)n \]
\[ \xi_i \xi_i: \ 4(p-2)n \quad m_i m_j: \ (p-1)n \]
\[ \xi_i \xi_j: \ 2(p-2)n \quad ij \]
\[ i \neq j \quad s_{ij}: \ \text{block; coefficients whose values are not needed} \]
\[ g_i m_i: \ 2(p-2)n \quad r_{ij}: \ \text{block; coefficients whose values are not needed} \]
\[ g_i m_j: \ (p-2)n \quad \]
\[ i \neq j \quad \text{All other elements: 0} \]

The following facts are obvious from consideration of the coefficients of the reduced equations:

1. \( \mu \) can be estimated independently of the other parameters,
2. \( \xi_i \) and \( m_j \) are not affected by \( s_{ij} \) and \( r_{ij} \) and can therefore be computed without regard to whether \( s_{ij} \) and \( r_{ij} = 0 \).
3. \( s_{ij} \) can be estimated entirely independently of all the other parameters, and
4. \( r_{ij} \) can be estimated entirely independently of all the other parameters.

In order to estimate \( \xi_i \) and \( m_j \), no matter what assumptions are made with respect to \( s_{ij} \) and \( r_{ij} \), only the \( g_i \) and \( m_j \) block of the above matrix needs to be considered. The inverse of this block has the following elements:

\[ g_i^i = \frac{(p-1)^2 / p^2(p-2)n}{(p-1)/p^2(p-2)n} \quad g_i^m = \frac{1}{p^2n} \]
\[ g_i^j = \frac{(p-1)^2 / p^2(p-2)n}{(p-1)/p^2(p-2)n} \quad m_i^i = \frac{2(p-1)/p^2n}{-2/p^2n} \]
\[ g_i^m = \frac{(p-1)^2 / p^2n}{(p-1)/p^2n} \quad m_i^j = \frac{1}{p^2n} \]

Since \( \hat{a} = \sum_{i=j} (y_{i..} + y_{j..}) g_i^i + \sum_j y_{i..} g_i^j \), and
\[ \hat{a} = \sum_{i=j} (y_{i..} + y_{j..}) m_i^i + \sum_j y_{i..} m_i^j, \]
- 84 -

\[ \hat{\mu} = \frac{1}{p(p-2)n} \left( (p-1) y_{a..} + y_{a...} - y_{..} \right), \]

\[ \hat{\sigma} = \frac{1}{pn} (y_{a..} - y_{a...}). \]

\[ \hat{\mu} \] is merely the mean of all the sample observations, that is \( \hat{\mu} = \frac{y_{a..}}{p(p-1)n} \)

Since the \( r_{ij} \) drop out of the \( s_{ij} \) equations,

\[ \hat{s}_{ab} = \frac{1}{2n} (y_{ab...} + y_{ba...}) - \hat{\mu} - \hat{\sigma}_a - \hat{\sigma}_b - \hat{m}_a/2 - \hat{m}_b/2 = \]

\[ \frac{1}{2(p-1)(p-2)n} \left( (p-1)(p-2) (y_{ab...} + y_{ba...}) - (p-1) (y_{a..} + y_{b..} + y_{a...} + y_{b...}) \right). \]

\[ \hat{r}_{ab} = \frac{1}{n} y_{ab...} - (\hat{\mu} + \hat{\sigma}_a + \hat{\sigma}_b + \hat{m}_a + \hat{m}_b) = \frac{1}{2pn} \left[ p(y_{ab...} - y_{ba...}) + y_{a..} - y_{b..} - y_{a...} + y_{b...} \right]. \]

The estimates of \( s_{ij} \) and \( r_{ij} \) can also be obtained by utilizing the elements of the inverse matrix. The inverse of the \( s_{ij} \) block has for its elements the following types of terms:

\[ s_{ij}s_{ij} = \frac{p-3}{2(p-1)n} \]

\[ s_{ia}s_{ib} = \frac{-(p-3)}{2(p-1)(p-2)n} \]

The inverse of the \( r_{ij} \) block has for its elements the following types of terms:

\[ r_{ab},r_{ba} = \frac{p-2}{2pn} \]

\[ r_{ab},r_{ac} = \frac{-1}{2pn} \]

Since the estimate of \( g_{ij} \) is different when \( m_j \) is assumed equal to zero than it is when \( m_j \) is assumed not equal to zero, the estimate under
the former assumption will now be given. The matrix of coefficients of
the least squares equations with respect to the $g_i$ block following elimina-
tion of the $g_p$ is

<table>
<thead>
<tr>
<th></th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>.........</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1$</td>
<td>4(p-2)n</td>
<td>2(p-2)n</td>
<td>.........</td>
</tr>
<tr>
<td>$g_2$</td>
<td>2(p-2)n</td>
<td>4(p-2)n</td>
<td>.........</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>.........</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>.........</td>
</tr>
</tbody>
</table>

The inverse of this matrix has the following types of elements,

$g_{i,i} = \frac{p-1}{2p(p-2)n}$

and

$g_{j,j} = \frac{-1}{2p(p-2)n}$

Then $\hat{\gamma}_a = \sum_i g_{a,i} y_{i,+} = \frac{1}{2p(p-2)n} \left[ p(y_{a,..} + y_{..,:}) - 2y_{..} \right].$

The $\hat{\gamma}_a$ is used to denote that this estimate is different from $\hat{\gamma}_1$, the
estimate when $m_j$ is not considered equal to zero.

If it is assumed that $g_1 = 0$, and that is the assumption necessary in
testing the hypothesis that $g_1 = 0$, the estimate of $m_j$ is different than
it is when $g_1$ is considered not equal to zero. In the former case $\hat{\gamma}_1 + \hat{m}_j =
\frac{y_{..,:}}{(p-1)n}$.

A somewhat different estimation problem with respect to the $r_{ij}$ arises
if the original model does not include $m_j$, as for example, in analyses of
poultry crosses. Then the restriction $\sum_i r_{ij} = 0$ for all $j$ is no longer a
sensible one. The only restriction now needed is $r_{ij} + r_{ji} = 0$. The lat-
ter restriction makes available $\frac{1}{2} p(p-1)$ degrees of freedom for the R sum
of squares rather than $\frac{1}{2} (p-1)(p-2)$ available in the original model. Consequently, the estimates of $r_{ij}$ are different in the two models. In the former $\hat{r}_{ab} = \frac{1}{n} y_{ab} - (\hat{\mu} + \hat{\varepsilon}_a + \hat{\varepsilon}_b + \hat{\varepsilon}_{ab})$, and in the latter $\check{r}_{ab} = \frac{1}{n} y_{ab} - (\hat{\mu} + \check{\varepsilon}_a + \check{\varepsilon}_b + \check{\varepsilon}_{ab}) = \frac{1}{2n} (y_{ab} - y_{ba})$. The $\check{}$ in place of $\hat{}$ denotes that the two estimates are different. The estimates of $\mu$ and $s_{ij}$ are the same in the two models. When $m_j$ is assumed to be non-existent, the elements of the block of the inverse matrix pertaining to $r_{ij}$ are $r_{ab}r_{ba} = 1/2n$, to $r_{ab}r_{ba} = -1/2n$, and to all others $= 0$.

b. Tests of hypotheses. The tests of hypotheses are easily effected since estimates of several of the sets of parameters are calculated independently of the others. The computational procedure for $m_j \neq 0$ will first be presented and then the computational procedure for $m_j = 0$.

(1) Hypothesis that $r_{ij} = 0$ assuming $m_j \neq 0$. Sum of squares
among $r_{ij} = \frac{1}{n} \sum_{ij} y_{ij}^2 - \frac{1}{2n} \sum_{ij} (y_{ij} + y_{ji})^2 + \sum_{i=j} (\check{\varepsilon}_i - \hat{\varepsilon}_i)
(y_{i..} + y_{..j}) - \sum_{j} y_{..j} \hat{m}_j$, degrees of freedom $= \frac{1}{2} (p-1)(p-2)$.

Error sums of squares $= \sum_{i,j,k} y_{ijk}^2 = \frac{1}{n} \sum_{ij} y_{ij}^2$, degrees of freedom $= p(p-1)(n-1)$.

(2) Hypothesis that $s_{ij} = 0$ assuming $m_j \neq 0$. Sum of squares
among $s_{ij} = \frac{1}{2n} \sum_{ij} (y_{ij} + y_{ji})^2 - \sum_{i=j} (y_{i..} + y_{..j}) \check{\varepsilon}_i
- \frac{1}{p(p-1)n} y_{..}^2$, degrees of freedom $= \frac{1}{2} p(p-3)$.

Error sum of squares (assuming $r_{ij} = 0$) $= \sum_{i,j,k} y_{ijk}^2 - \frac{1}{2n} \sum_{ij} (y_{ij} + y_{ji})^2 + \sum_{i=j} (y_{i..} + y_{..j})(\check{\varepsilon}_i - \hat{\varepsilon}_i) - \sum_{j} y_{..j} \check{m}_j$, degrees of freedom $= \frac{1}{2} (p-1)(2pn - p - 2)$. 
\( \hat{e}_i \) denotes estimates of \( e_i \) when \( m_j \neq 0 \).

\( \hat{e}_i \) denotes estimates of \( e_i \) when \( m_j = 0 \).

The error sum of squares, assuming \( r_{ij} \neq 0 \), is the same as in the test of hypothesis that \( r_{ij} = 0 \).

(3) **Hypothesis that \( m_j = 0 \).** Sum of squares among \( m_j = \sum_{ij} (y_{ij} - \hat{\mu}_j - \hat{\xi}_i)^2, \) degrees of freedom = \( p-1 \).

Error sum of squares (assuming \( r_{ij} = \xi_{ij} = 0 \)) = \( \sum_{ijk} y_{ijk}^2 - \sum_{i,j} (y_{ij} - \hat{\xi}_i) \hat{\xi}_i - \sum_j y_{.j} \hat{\mu}_j, \) degrees of freedom = \( np(p-1) - 2p + 1 \).

Error sum of squares, assuming \( r_{ij} = 0, \xi_{ij} \neq 0 \), is the same as the error sum of squares for the test of the hypothesis that \( \xi_{ij} = 0 \), assuming \( r_{ij} = 0 \). The error sum of squares assuming \( r_{ij}, \xi_{ij} \neq 0 \) is the same as the error sum of squares for the test of hypothesis that \( r_{ij} = 0 \).

(4) **Hypothesis that \( \xi_i = 0 \) assuming \( m_j \neq 0 \).** Sum of squares among \( \hat{\xi}_i = \sum_{ij} (y_{ij} - \hat{\mu}_j - \hat{\hat{\xi}}_i)^2 \hat{\xi}_i - \frac{1}{(p-1)n} \sum_j y_{.j}^2 + \frac{1}{p(p-1)n} y^2 \ldots, \) degrees of freedom = \( p-1 \).

The error sum of squares is the same as in the test of the hypothesis that \( m_j = 0 \).

(5) **Hypothesis that \( r_{ij} = 0 \) assuming \( m_j = 0 \).** The sum of squares among \( r_{ij} = \frac{1}{n} \sum_{ij} y_{ij}^2 - \frac{1}{2n} \sum_{ij} (y_{ij} - \hat{y}_{ij})^2, \) degrees of freedom = \( \frac{1}{2} p(p-1) \).

Error sum of squares = \( \sum_{ijk} y_{ijk}^2 - \frac{1}{n} \sum_{ij} y_{ij}^2, \) degrees of
(6) **Hypothesis that \( s_{ij} = 0 \) assuming \( m_j = 0 \).** Sum of squares among \( s_{ij} = \frac{1}{2n} \sum_{ij} (y_{ij} - y_{ij.})^2 - \sum_{i=j} (y_{i..} - y_{..j}) \hat{e}_i 
\quad - \frac{1}{p(p-1)n} y^2 ...$, degrees of freedom = \( \frac{1}{2} p(p-3) \).

Error sum of squares (assuming \( r_{ij} = 0 \)) = \( \sum y_{ijk}^2 \n\quad - \sum_{i=j} (y_{i..} - y_{..j}) \hat{e}_i \), degrees of freedom = \( \frac{1}{2} p(p-1)(2n-1) \).

The error sum of squares assuming \( r_{ij} \neq 0 \) is the same as is the test of the hypothesis that \( r_{ij} = 0 \), assuming \( m_j \neq 0 \).

(7) **Hypothesis that \( g_i = 0 \) assuming \( m_j = 0 \).** Sum of squares among \( g_i = \sum_{i=j} (y_{i..} - y_{..j}) \hat{e}_i \), degrees of freedom = \( p-1 \).

Error sum of squares (assuming \( s_{ij} = r_{ij} = 0 \)) = \( \sum y_{ijk}^2 \n\quad - \sum_{i=j} (y_{i..} - y_{..j}) \hat{e}_i \), degrees of freedom = \( (p-1)(2n-1) \).

The error sum of squares (assuming \( s_{ij} \neq 0, r_{ij} = 0 \)) is the same as the error sum of squares for the test of \( s_{ij} = 0 \) under the assumption that \( r_{ij} = 0 \).

Error sum of squares (assuming \( s_{ij}, r_{ij} \neq 0 \)) is the same as in the test of the hypothesis that \( r_{ij} = 0 \).

c. **Estimation of components of variances.** The estimates of \( \sigma^2_g \), \( \sigma^2_m \), \( \sigma^2_s \), \( \sigma^2_r \), and \( \sigma^2_e \) are easily computed by equating the sums of squares for \( R, S, M, G \), and error to the expected values of these sums of squares. Assuming that \( m_j \neq 0 \), the coefficients of the variances in the different expectations are as follows:
d. Adjustment of least squares estimates. Each of the least squares estimates can easily be written in terms of the parameters, thus,

\[ \hat{E}_{e_a} = \frac{p-1}{p} \sum_{i \neq a} g_i + \frac{1}{p} \sum_{j \neq a} (s_{a_j} + r_{a_j}) - \frac{1}{p(p-2)} \sum_{i \neq j \neq a} (2s_{ij} + r_{ij}) + \text{errors, when } m_j \neq 0 \]

\[ \hat{E}_{e_a} = \frac{p-1}{p} \sum_{i \neq a} g_i + \frac{1}{p} \sum_{j \neq a} a_{aj} - \frac{2}{p(p-2)} \sum_{i \neq j \neq a} s_{ij} + \frac{1}{2p} \sum_{j \neq a} r_{ia} + \text{errors, when } m_j = 0 \]

\[ \hat{E}_{e_a} = \frac{p-1}{p} m_a + \frac{1}{p} \sum_{j \neq a} r_{ia} - \frac{1}{p(p-2)} \sum_{j \neq a} r_{aj} + \text{errors} \]

\[ \hat{E}_{e_{ab}} = \frac{p-3}{p-1} \sum_{i \neq a, b} s_{ab} - \frac{p-3}{(p-1)(p-2)} \sum_{i \neq j \neq a, b} (s_{a_j} + s_{b_j}) + \frac{2}{(p-1)(p-2)} \sum_{i \neq j \neq a, b} s_{ij} + \frac{2}{2(p-1)(p-2)} \sum_{i \neq j \neq a, b} (r_{ab} + r_{ba}) - \frac{p-3}{2(p-1)(p-2)} \sum_{i \neq j \neq a, b} (r_{aj} + r_{bj}) + \frac{1}{(p-1)(p-2)} \sum_{i \neq j \neq a, b} r_{ij} + \text{errors, when } m_j \neq 0 \]

\[ \hat{E}_{e_{ab}} = \frac{p-2}{2p} (r_{ab} - r_{ba}) + \frac{1}{2p} \sum_{i \neq j \neq a, b} (r_{bj} - r_{aj}) + \frac{1}{2p} \sum_{i \neq j \neq a, b} (r_{ia} - r_{ib}) + \text{errors, when } m_j \neq 0 \]

Then since \( b(\hat{e}_e) = \frac{E}{E^2} \hat{e}_e \), the regressions to be applied to the various
least squares estimates are as follows:

\[ b(\hat{e}_{i1}) = \frac{\sigma^2}{\sigma^2 + \frac{\sigma^2}{p-2} + \frac{(p-1) \sigma^2}{p(p-2)} + \frac{(p-1) \sigma^2}{p(p-2)n}} \text{ when } m_j \neq 0 \]

\[ b(\hat{e}_{i1}) = \frac{\sigma^2}{\sigma^2 + \frac{\sigma^2}{p-2} + \frac{\sigma^2}{2(p-2)} + \frac{\sigma^2}{2(p-2)n}} \text{ when } m_j = 0 \]

\[ b(m_j \hat{e}_{ij}) = \frac{\sigma^2}{\sigma^2 + \frac{\sigma^2}{p} + \frac{\sigma^2}{pn}} \]

\[ b(s_{ij} \hat{s}_{ij}) = \frac{\sigma^2}{\sigma^2 + \frac{\sigma^2}{2} + \frac{\sigma^2}{2n}} \text{ when } m_j \neq 0 \text{ and when } m_j = 0 \]

\[ b(r_{ij} \hat{r}_{ij}) = \frac{\sigma^2}{\sigma^2 + \frac{\sigma^2}{n}} \text{ when } m_j \neq 0 \text{ and when } m_j = 0 \]

An attempt was made to write general formulas for predicting the results of specific crosses from the least squares estimates obtained from a balanced design, but the algebra was too complicated. For particular numerical values of \( p, n, \) and the variances, the necessary equations are easy to solve.

2. Top cross test

Assume that \( q \) different breeds or strains are used as tester females, that there are \( p \) inbred lines to be tested, and that \( n \) units are tested of each of the possible crosses of inbred males on tester females. A reasonable
mathematical model is

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \]

where \( y_{ijk} \) denotes the \( k \)th unit of the cross between the \( i \)th inbred lines by the \( j \)th breed, \( \alpha_i \) is an effect common to progeny of the \( i \)th inbred line (genetically it corresponds to the \( g_1 \) of the single cross design previously described), \( \beta_j \) is an effect common to progeny of the \( j \)th breed, and \( \gamma_{ij} \) is an effect common to progeny of the cross of the \( i \)th line by the \( j \)th breed.

Data from such an experiment can be treated by the simple analysis appropriate to a two-way factorial classification with equal subclass numbers. The tests of hypotheses come from a straightforward analysis of variance, thus,

- Among \( G \): \( p-1 \) d.f.
- Among \( B \): \( q-1 \) d.f.
- \( G \times B \): \( (p-1)(q-1) \) d.f.
- Error: \( pq(n-1) \) d.f.

The parameter estimates are easily obtained as follows:

\[ \hat{\alpha} = \frac{\bar{y}_{.. \cdot}}{pn} \]
\[ \hat{\beta}_j = \frac{\bar{y}_{j.. \cdot} - \bar{y}_{.. \cdot}}{pn} - \hat{\mu} \]
\[ \hat{\gamma}_{ij} = \frac{\bar{y}_{ij..} - \bar{y}_{.. \cdot}}{n} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j \]

The variances and covariances for the \( \hat{\alpha}_i \) are,

\[ V(\hat{\alpha}_i) = \frac{p-1}{pn} \sigma^2 \]
\[ CV(\hat{\alpha}_i \hat{\alpha}_j) = -\frac{1}{pqn} \hat{\sigma}^2 \]

Therefore, \( V(\hat{\alpha}_i - \hat{\alpha}_j) = \frac{2}{qn} \hat{\sigma}^2 \).
3. Balanced incomplete block design for farm testing

Assume that there are \( p \) lines to be tested, that two lines are compared per farm, that each line is compared with each of the other lines once and only once, and that \( n \) units are tested per line per farm. Assume also that, within farms, lines are orthogonal to breed of dam, age, and other extraneous factors. \( \frac{1}{2} p(p-2) \) farms are required for this design.

The mathematical model is

\[
y_{ijk} = \mu + e_i + f_j + g_{ij} + e_{ijk}
\]

\( y_{ijk} \) denotes the \( i^{th} \) unit of the \( i^{th} \) line on the \( j^{th} \) farm. \( e_i \) is an effect common to progeny of the \( i^{th} \) line. Genetically it is presumably the same as the \( e_i \) in previous designs. \( f_j \) is an effect common to progeny of the \( j^{th} \) farm. It includes both genetic and environmental factors. \( g_{ij} \) is an effect common to progeny of the \( i^{th} \) line produced on the \( j^{th} \) farm.

The analysis is easily carried out as described below.

\[ \hat{\varepsilon}_i = \frac{1}{pn} (2y_{i..} - \text{sum of observations over all farms on which the } i^{th} \text{ line was tested}) \]

\[ \hat{\mu} + \hat{f}_j = \frac{1}{2n} (y_{..j} - n \hat{e}_a - n \hat{e}_b) \]

where \( a \) and \( b \) denote the two lines tested on the \( j^{th} \) farm.

\( G \times F \) sum of squares = \( R(\mu, g, f, gf) - R(\mu, g, f) \); degrees of freedom = \( p(p-1)(p-2)/2 \).

\[ R(\mu, g, f, gf) = \frac{1}{n} \sum_{ij} y_{ij}^2 \]

\[ R(\mu, g, f) = \sum_i e_i y_{i..} + \sum_j (\hat{\mu} + \hat{f}_j) y_{..j} \]

Error sum of squares = \( R(T) - R(\mu, g, f, gf) \); degrees of freedom = \( p(p-2)(n-1) \).

\( F \) sum of squares = \( R(\mu, g, f) - R(\mu, g) \); degrees of freedom = \( \frac{1}{2} (p + 1) \)
(p-2). \( R(\mu, g) = \frac{1}{(p-1)n} \sum_i y_i^2 \).

\( G \) sum of squares = \( R(\mu, g, f) - R(\mu, f) \); degrees of freedom = \( p-1 \).

\( R(\mu, f) = \frac{1}{2n} \sum_j y_j^2 \).

The sampling errors are

\( V(\hat{\varepsilon}_i) = \frac{2(p-1)}{p^2n} \sigma_e^2 \)

\( CV(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = \frac{2}{pn} \sigma_e^2 \)

\( V(\hat{\varepsilon}_i - \hat{\varepsilon}_j) = \frac{4}{pn} \sigma_e^2 \).

4. Three-way crosses

From the standpoint of commercial utilization the three-way cross is of greater potential importance than the two-way cross. The analysis of the three-way cross is more complex than is the analysis of the single cross, but the computational procedure for a balanced three-way cross is not difficult. In this balanced design it is assumed that inbred males are mated with single cross females and that \( n \) units of each of the possible three-way crosses are tested. The way in which the single cross females are formed is ignored; that is, reciprocal crosses are considered to be the same with respect to performance as dams.

The mathematical model is

\[ y_{ijkl} = \mu + \delta_i + \frac{1}{2} \varepsilon_j + \frac{1}{2} \varepsilon_k + \frac{1}{2} m_j + \frac{1}{2} m_k + s_{jk} + t_{ijk} + e_{ijkl}. \]

This model was described in Section II H 4. The least squares equations for this design are as follows:
\[ \mu = \frac{1}{2} np(1)(p-2) \mu + n(p-1)(p-2) \sum_i g_i + \frac{1}{2} n(p-1)(p-2) \sum_j m_j \]
\[ = n(p-2) \sum_{j,k} t_{ijk} + n \sum_{i,j,k} s_{ijk} \]
\[ = n(p-1)(p-2) \mu + \frac{3}{4} n(p-1)(p-2) g_a + \frac{5}{4} n(p-2) \sum_{i,j,k} g_i + \frac{1}{2} n(p-1) \]
\[ (p-2) m_a + \frac{3}{4} n(p-2) \sum_{j,k} m_j + n \sum_{j,k} s_{jk} + \frac{1}{2} (p-2) n \sum_k s_{ak} \]
\[ + n \sum_{j,k} t_{ijk} + n \sum_{i,k} t_{iak} = y_a \ldots + \frac{1}{2} y_a \ldots \]
\[ = n(p-1)(p-2) \mu + \frac{1}{2} n(p-1)(p-2) g_a + \frac{3}{4} n(p-2) \sum_{i,j,k} g_i + \frac{1}{2} n \]
\[ (p-1)(p-2) m_a + \frac{1}{2} (p-2) n \sum_{j,k} m_j + \frac{1}{2} (p-2) n \sum_{k} s_{ak} + \frac{1}{2} n \]
\[ \sum_{i,k} t_{iak} = \frac{1}{2} y_a \ldots \]
\[ = n(p-2)(\mu + \frac{1}{2} g_a + \frac{1}{2} g_b) + n \sum_{i,j,k} g_i + n(p-2)(\frac{1}{2} m_a + \frac{1}{2} m_b + \frac{1}{2} m_c + s_{bc} + t_{abc}) = y_{abc} \]

By reducing the above equations to the number of independent parameters, it can be shown that each set of parameters is estimated independently of the others except for the \(g_i\) and \(m_j\) with one another. Consequently, \(\hat{g}_i, \hat{m}_j, V(\hat{g}_i), V(\hat{m}_j)\), and tests of the hypotheses that \(g_i = 0\) and that \(m_j = 0\) are independent of what assumptions are made with respect to \(s_{jk}\) and \(t_{ijk}\).

Similarly, \(\hat{s}_{jk}, V(\hat{s}_{jk})\), and the test of the hypothesis that \(s_{jk} = 0\) are independent of whether or not the \(t_{ijk}\) are assumed equal to zero. The inverse elements of the \(g_i, m_j\) block are

\[ g_i^\dagger g_i = \frac{2(p-1)}{p^2(p-3)n} \]
\[
g^i_{
abla j} = -\frac{2}{p^2(p-3)n}
\]
\[
g^i_{mj} = -\frac{2(p-1)(p-4)}{p^2(p-2)(p-3)n}
\]
\[
g^i_{mj} = \frac{2(p-4)}{p^2(p-2)(p-3)n}
\]
\[
m^i_{mj} = \frac{2(p-1)(3p-8)}{p^2(p-2)(p-3)n}
\]
\[
m^i_{mj} = -\frac{2(3p-8)}{p^2(p-2)(p-3)n}
\]

The least squares estimates are

\[
\hat{\mu} = \frac{2}{np(p-1)(p-2)} y_1...
\]
\[
\hat{\sigma}_a = \frac{2}{p(p-2)(p-3)n} \left[ (p-2) y_a... + y.a.. - y_1... \right]
\]
\[
\hat{\sigma}_a = \frac{2}{p(p-2)(p-3)n} \left[ -(p-2) y_a... + (p-2) y.a.. - y_1... \right]
\]
\[
\hat{\sigma}_{ab} = \frac{1}{(p-1)(p-2)^2n} \left[ (p-1)(p-2) y_{ab..} - (p-1)(y.a.. + y.b..) + 2y_1... \right]
\]
\[
\hat{\sigma}_{abc} = \frac{1}{n} y_{abc..} - (\hat{\mu} + \hat{\sigma}_a + \frac{1}{2} \hat{\sigma}_b + \frac{1}{2} \hat{\sigma}_c + \frac{1}{2} \hat{\sigma}_{ab} + \frac{1}{2} \hat{\sigma}_{ac} + \frac{1}{2} \hat{\sigma}_{bc}).
\]

Under the assumption that \( m_j = 0 \), the estimates of \( g_1 \) are different from those given above. When \( m_j = 0 \), they are denoted by \( \hat{\gamma}_1 \). The elements of the inverse matrix with respect to \( \hat{\gamma}_1 \) are

\[
\hat{g}^i_{\hat{g}^j} = \frac{4(p-1)}{p(p-2)(3p-8)n}
\]
\[
\hat{g}^i_{\hat{g}^j} = \frac{-4}{p(p-2)(3p-8)n}
\]
Then \( \hat{\mathbf{x}}_1 = \frac{2}{p(p-2)(3p-8)n} \left[ p(2y_{1\ldots} + y_{1\ldots}) - 4y_{1\ldots} \right] \).

Under the assumption that \( g_1 = 0 \), the elements of the inverse matrix with respect to \( \mathbf{y}_j \) are

\[
m^{\prime \prime}m^j = \frac{4(p-1)}{p(p-2)^2n} \\
m^{\prime \prime}m^{j'} = -\frac{4}{p(p-2)^2n}
\]

The estimates of \( m_j \) under the assumption that \( g_1 = 0 \) are \( \hat{m}_j = \frac{1}{p(p-2)^2n} \) \((py_{1\ldots} - 2y_{1\ldots})\).

The sums of squares for the various tests of hypotheses are as follows:

- **T sum of squares**
  \[
  \text{T sum of squares} = \frac{1}{n} \sum y_{1jk}^2 - \frac{1}{2} \sum_j (\hat{\mathbf{y}}_j - \hat{\mathbf{y}}_{1\ldots}) y_{1\ldots} - \frac{1}{i=1} (y_{1\ldots} \cdot \frac{1}{2} y_{1\ldots}) \\
  \]  
  degrees of freedom = \( \frac{1}{2} (p-1) \)

- **S sum of squares**
  \[
  \text{S sum of squares} = \frac{1}{n(p-2)} \sum_{jk} y_{1jk}^2 - \frac{1}{2} \sum_j \hat{m}_j y_{1\ldots} - \frac{np(p-1)}{(p-2)} y_{1\ldots}^2 \\
  \]  
  degrees of freedom = \( \frac{1}{2} p(p-3) \)

- **M sum of squares**
  \[
  \text{M sum of squares} = \sum_{i=j} (\hat{E}_i - \hat{E}_1) (y_{i\ldots} \cdot \frac{1}{2} y_{1\ldots}) + \frac{1}{2} \sum_j \hat{m}_j y_{1\ldots} \\
  \]  
  degrees of freedom = \( p-1 \)

- **G sum of squares (assuming \( m_j \neq 0 \))**
  \[
  \text{G sum of squares} = \sum_{i=j} \hat{E}_i (y_{i\ldots} \cdot \frac{1}{2} y_{1\ldots}) + \frac{1}{2} \sum_j (\hat{m}_j - \frac{1}{2} \hat{m}_j) y_{1\ldots} \\
  \]  
  degrees of freedom = \( p-1 \)

- **G sum of squares (assuming \( m_j = 0 \))**
  \[
  \text{G sum of squares} = \sum_{i=j} \hat{E}_i (y_{i\ldots} \cdot \frac{1}{2} y_{1\ldots}) \\
  \]  

- **Error sum of squares (assuming \( g_1, m_j, \hat{m}_j, \hat{e}_{ijk}, t_{ijk} \neq 0 \))**
  \[
  \text{Error sum of squares} = \sum_{ijkl} y_{ijkl}^2 - \frac{1}{n} \sum_{ijk} y_{1jk}^2 
  \]
Error sum of squares (assuming $t_{ijk} = 0$) = $\sum y_{ijk}^2 - \frac{1}{n(p-2)} \sum_{jk} y_{ijk}^2$

$$y_{ijk}^2 + \frac{1}{2} \sum_{j} m_{j} y_{..j}$$

Error sum of squares (assuming $s_{jk} = t_{ijk} = 0$) = $\sum y_{ijk}^2 - \sum_{i=j} \hat{e}_i$

$$(y_{..i} + \frac{1}{2} y_{..j}) - \frac{1}{2} \sum_{j} m_{j} y_{..j} = \frac{2}{np(p-1)(p-2)} y^2$$

Error sum of squares (assuming $m_{j} = s_{jk} = t_{ijk} = 0$) = $\sum y_{ijk}^2 - \sum_{i=j} \hat{e}_i$

$$\hat{e}_i (y_{..i} + \frac{1}{2} y_{..j}) - \frac{2}{np(p-1)(p-2)} y^2$$

5. **Appraisal of lines by their own performances**

Obviously, one of the easiest ways to test lines is to measure their own performances. Let the model in this case be,

$$y_{ij} = \mu + 2g_i + m_i + e_{ij}$$

$y_{ij}$ denotes the $j^{th}$ unit of the $i^{th}$ line. $g_i$ is the additive genetic effect of the line the same as is $g_1$ in the other models. $m_i$ is the maternal ability of the line.

It is clear that $g_i$ and $m_i$ are completely confounded if the young are raised by dams of their own line. The post-natal portion of the maternal effect can, however, be separated if at birth some of the progeny are switched to other lines. Assume that there are $p$ lines to test and that at birth the young are switched in such a way that $n$ units of each line are raised by all of the lines. Then the analysis described below will separate $g_i$ and $m_i$. The mathematical model now needs to be modified as follows:

$$y_{ijk} = \mu + 2g_i + m_j + g_{mij} + e_{ijk}$$
\[ y_{ijk} \] denotes the \( k^{\text{th}} \) unit of the \( i^{\text{th}} \) line raised by dams of the \( j^{\text{th}} \) line. \( g_i \) denotes an effect common to progeny of the \( i^{\text{th}} \) line. It includes genetic and pre-natal maternal effects. \( m_j \) denotes an effect common to young raised by dams of the \( j^{\text{th}} \) line. \( g_{mij} \) is an effect common to progeny of the \( i^{\text{th}} \) line reared by \( j^{\text{th}} \) line females.

The analysis is carried out in the same manner as is any two-way classification with equal subclass numbers.

6. **Relative efficiencies of different balanced designs**

The single cross test yields information on general combining abilities of the lines, maternal abilities of the lines, specific combining abilities of the lines with respect to one another, and sex linkage effects. Nothing is learned concerning the maternal performance of different single crosses. The three-way cross yields information on general combining ability, general combining ability with respect to mothering ability, specific combining ability with respect to performance as single cross females, three-way interaction, and if desired, the sex linkage effect with respect to performance as line of dam. Top cross and farm tests yield information on the general combining abilities of the lines and something on the specific combining ability with respect to crosses of specific lines with specific breeds or lines making up the tester stock. No information is obtained concerning maternal ability of the lines or crosses among the lines, and no information is obtained concerning specific combining ability among the tested inbred lines unless the inbred lines are used as tester females.

Since all of these tests provide information on general combining ability, it is of interest to see how they compare in the amount of such
information. If it is assumed that $\sigma_e^2 = 1$ for all of the designs (an assumption of equal $\sigma_e^2$ is not strictly correct, but environment makes up such a large portion of $\sigma_e^2$ that little inaccuracy results from the assumption) and that the total number of experimental units $= pk$, the variance of the difference between the estimates of general combining ability of two lines for each of the different types of tests is as follows:

Own performance $= 1/2k$

Top cross $= 2/k$

Farm test $= 4(p-1)/pk$

Single cross ($m_j \neq 0$) $= 2(p-1)^2/p(p-2)k$

Single cross ($m_j = 0$) $= (p-1)/(p-2)k$

3-way cross ($m_j \neq 0$) $= 2(p-1)(p-2)/p(p-3)k$

3-way cross ($m_j = 0$) $= 4(p-1)/(3p-8)k$

Table 1 is presented below for the purpose of comparing these variances for different values of $p$.

<table>
<thead>
<tr>
<th>$p$</th>
<th>Own Performance</th>
<th>Top Cross</th>
<th>Farm Test</th>
<th>Single Cross $m_j \neq 0$</th>
<th>Single Cross $m_j = 0$</th>
<th>3-Way Cross $m_j \neq 0$</th>
<th>3-Way Cross $m_j = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>.50</td>
<td>2.00</td>
<td>3.00</td>
<td>2.25</td>
<td>1.50</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>6</td>
<td>.50</td>
<td>2.00</td>
<td>3.33</td>
<td>2.06</td>
<td>1.25</td>
<td>2.22</td>
<td>2.00</td>
</tr>
<tr>
<td>8</td>
<td>.50</td>
<td>2.00</td>
<td>3.50</td>
<td>2.04</td>
<td>1.17</td>
<td>2.10</td>
<td>1.75</td>
</tr>
<tr>
<td>10</td>
<td>.50</td>
<td>2.00</td>
<td>3.60</td>
<td>2.02</td>
<td>1.12</td>
<td>2.06</td>
<td>1.64</td>
</tr>
<tr>
<td>20</td>
<td>.50</td>
<td>2.00</td>
<td>3.80</td>
<td>2.01</td>
<td>1.06</td>
<td>2.01</td>
<td>1.46</td>
</tr>
<tr>
<td>30</td>
<td>.50</td>
<td>2.00</td>
<td>3.87</td>
<td>2.00</td>
<td>1.04</td>
<td>2.00</td>
<td>1.41</td>
</tr>
<tr>
<td>100</td>
<td>.50</td>
<td>2.00</td>
<td>3.96</td>
<td>2.00</td>
<td>1.01</td>
<td>2.00</td>
<td>1.36</td>
</tr>
<tr>
<td>$\infty$</td>
<td>.50</td>
<td>2.00</td>
<td>4.00</td>
<td>2.00</td>
<td>1.00</td>
<td>2.00</td>
<td>1.33</td>
</tr>
</tbody>
</table>
It can be seen that, under the hypothesis that \( m_j = 0 \), the information on general combining ability furnished by the single cross test approaches twice that of a top cross test, and the information supplied by the three-way cross approaches \( 1\frac{1}{2} \) times that of the top cross test. When \( m_j \neq 0 \), both cross line tests rapidly approach the top cross test in efficiency. The farm test becomes relatively less and less efficient as the number of lines increases. This disadvantage should, of course, be balanced against the potentially much larger numbers which can be included in this type of experiment.

J. Estimation of Optimum Number of Lines to Test in Single Crosses

Assuming that the limiting factor in a program of making and testing inbred lines in single crosses is the number of units which can be tested and assuming that good estimates of the relative sizes of \( \sigma^2_g \), \( \sigma^2_m \), \( \sigma^2_s \), \( \sigma^2_r \), and \( \sigma^2_e \) are available, the question as to the number of lines to test in a balanced single cross design in order to maximize the probability of selecting, on the basis of the test, the line with highest \( g_i \), the line with highest \( g_i + m_i \), and the single cross with the highest \( g_i + g_j + m_i + m_j + s_{ij} + r_{ij} \) can be answered. The method involves the following steps:

1. Constructing the appropriate index as described in Section II G 6.
2. Computing the variance of the index.
3. Expressing the variance of the index as a function of the number of lines tested (\( p \)), the total number of experimental units, and the numerical values of the different variances.
4. Choosing a value of \( p \) such that the expected value of the largest
deviate in a sample of size \( n' \) from a normal population with mean
= 0 and variance = \( \sigma(I) \) is a maximum. \( n' = p \) if selection is for
\( g_i \) or for \( g_i + m_i \), and \( n' = \frac{1}{2} p(p-1) \) if selection is for \( g_i + g_j + m_j + s_{ij} + r_{ij} \).

The methods for constructing the appropriate indexes were described in
Section II G 6. The expected values needed for constructing the indexes
can be derived from the expected values of the least squares estimates pre-
sented in Section II I 1 d. For example, if it is desired to select for
general combining ability on the basis of a single cross test, and it is
assumed that \( \sigma_m^2 = \sigma_r^2 = 0 \), the appropriate index is

\[
I = \frac{E g_i g_i}{E g_i^2}
\]

\[
V(I) = \frac{(E g_i g_i)^2}{(E g_i^2)^2} \quad E g_i^2 = \frac{(E g_i g_i)^2}{E g_i^2}
\]

\[
= \frac{p-1}{p} \left( \frac{\frac{\sigma_i^2}{p} + \frac{\sigma_m^2}{p-2} + \frac{\sigma_r^2}{2(p-2)n}}{\sigma_i^4} \right)
\]

Let the total number of experimental units = \( k = p(p-1)n \). Then \( n = \frac{k}{p(p-1)} \).
Substituting the numerical values of the variances and of \( k \) in the formula
for the variance of the index, \( V(I) \) becomes a function of \( p \), the number of
lines tested. Then by a process of trial and error, the value of \( p \) which
maximizes \( \frac{\sigma_i^2}{p} \) can be obtained, \( \bar{x}_p \) being the expected value of the lar-
gest deviate in a sample of size \( p \) from a normal population with mean = 0
and variance = 1.
III. ANALYSIS OF EXPERIMENTAL RESULTS

A. The Source and Scope of the Data

The data analyzed in this study come from crosses among the twelve inbred Poland China lines developed at the Iowa Agricultural Experiment Station. The single cross litters were farrowed during the fall seasons from 1942 to 1947, inclusive. The first of the inbred lines was begun in 1930 as a four sire line. Six of the lines were started in 1937 as selections from the original four sire line. The remaining six lines were established in 1937 and 1938. Four of these latter six lines are unrelated to each other or to any of the earlier six lines, while two were formed by crossing unrelated lines selected from the other ten. In addition to the four sire line, three of the lines have been maintained as two sire lines, and eight have been maintained as one sire lines. Vernon (1948) has reported in detail the history of each of the twelve lines. The average inbreeding coefficients of the sows of the different lines at the time they were used in these crossing experiments are given in Table 8.

Beginning with the fall of 1942 as many cross-line litters as possible were produced, the number depending upon the physical facilities and the availability of females above the requirements for maintaining the lines. The production of single crosses had several purposes, namely,

1. To compare the performance of single cross pigs with that of inbred pigs

2. To produce single cross sows for use in crossing with a third line

3. To measure the general combining ability of the various lines in crosses with other lines
4. To obtain some information concerning the value of specific crosses.

Although it was not possible to produce all of the 132 different possible reciprocal crosses in any one year, it was planned to eventually test a few litters of each of such crosses. The general plan was to use, if possible, two sires from each line in a particular season and to cross each of these sires with a sow from each of the other lines to be tested that season. After weaning the pigs were grouped according to age and fed in large lots containing a little late fall pasture. The particular cross had no bearing on the group in which the individual pig was fed. Feeding and management practices were as nearly alike from year to year as it was possible to make them. The following data were obtained for all of the litters:

1. Number of pigs, dead or alive, at birth
2. Weight of the individual pig at birth
3. Weights of individual pigs at 21 days, 56 days, and 154 days.

If the weighing was not done at exactly these ages, corrections were applied according to the formulas developed by Whatley (1937). One weight only was taken for each of the first three ages, but in the case of 154 day weights one weight was obtained a short time prior and a second weight a short time subsequent to 154 days, and the two adjusted weights were averaged to obtain the 154 day weight for each pig.

In Table 2 is presented a two-way classification of line of sire by line of dam with respect to number of litters produced during the six year test period. This table vividly illustrates the difficulty in obtaining
a balanced design for testing when it is necessary at the same time to maintain lines. The number of litters obtained for different crosses varied from 0 to 12 and it was possible to test only 77 of the 132 possible crosses during a six year period. The large number of missing subclasses and the unequal numbers in the subclasses illustrate why a least squares solution was needed to obtain satisfactory estimates of the line characteristics.

From the standpoint of estimating specific combining ability the total number of crosses, ignoring which way the cross was made, is the important consideration. This is particularly true if maternal effect is small or entirely lacking. In Table 3 the number of litters of each of the crosses, ignoring the way in which the cross was made, is presented.
Table 3

Number of Litters Tested of Each of the Possible
Crosses Ignoring Which Way the Cross Was Made

<table>
<thead>
<tr>
<th>Line</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>S</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>C</td>
<td>11</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>D</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td>57</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>0</td>
<td>3</td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It will be noted that the number of litters tested for each of the crosses varied from 0 to 17, and that 50 of the possible 66 crosses were tested.

B. Methods of Analysis Pertaining to the Available Single Cross Data

Methods for analyzing single cross data were described in detail in Section II, but there are some additional problems to be discussed in connection with this particular set of data. One of the first questions which arose was what measurement or measurements should be used to appraise the lines and the line crosses. For example, data were available on litter numbers and litter weights at various ages and individual weights at different ages. It was decided for purposes of this study that the maximum amount of information could be obtained for a given amount of computation by considering the litter as the experimental unit. The computations are extremely laborious at best, so it seemed desirable to work with different
measurements on the same experimental unit in order that the coefficients in the least squares equations, the elements of the inverse matrices, and the expected values of the different reductions in sums of squares would be the same for each measurement. The computations required to analyze thoroughly the data with respect to all of the eight different measurements on litters are less than those required to analyze only one criterion for litters and one for individual pig weight. Furthermore, from an economic standpoint, the weight of the litter at near market age is one of the most important single criteria.

It was also necessary to decide what sources of variation in addition to line or cross should be taken into account in the analysis. Some of the extraneous factors which might be thought important are sex ratio, age of sow, inbreeding of the sow, inbreeding of the litter, year in which the litter was farrowed, and time during a particular season when the litter was farrowed. The possibility of using correction factors was soon abandoned since there appeared to be none available which one could be sure was applicable to the particular data at hand. It was therefore a question as to just how many additional criteria of classification in addition to lines and crosses could be handled in a least squares solution. Following an inspection of the data, sex ratio was discarded as an important source of variability, particularly since it was hardly conceivable that differences in sex ratio among lines and crosses would be large enough to bias the appraisal of the lines and crosses. The coefficients of inbreeding of the cross line pigs were too small to have had much effect upon the performance of the pigs. Inbreeding of the cross-line pigs therefore was abandoned as a criterion of classification in the analysis. The age and inbreeding of
the sow and the year in which the litters were born seemed important enough
to consider in the analysis. A preliminary analysis of 154 day litter
weights, however, showed that the differences among years were much too
small to merit further consideration. The test of hypothesis that years =
0 resulted in the following analysis of variance:

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among years</td>
<td>5</td>
<td>63,145</td>
</tr>
<tr>
<td>Error</td>
<td>201</td>
<td>93,522</td>
</tr>
</tbody>
</table>

Then an inspection of the marginal means with respect to years for each
of the eight litter measurements made it very plain that the tests of sig-
nificance of yearly differences for the other seven criteria of classifi-
cation in addition to 154 day weight would result in mean squares for
years no larger than mean squares for error. The fact that yearly differ­
ences are of no consequence in this set of data is very fortunate from the
standpoint of appraising the lines, since the distribution of lines tested
from year to year was quite unbalanced. (Tables 4 and 5) Consequently,
the sampling errors of the estimates are rather materially less than they
would be if it were necessary to consider yearly differences. It should be
pointed out that differences among years are almost completely confounded
with annual changes in the genetic merit of the lines since there was little
carry-over of sows or bears from one year to the next.

A preliminary analysis of 154 day weights and an inspection of the mar-
ginal means showed that age of sow accounted for a significant amount of
variation among litters. Since it is almost certain that the regression of
litter number and weight on age of sow is curvilinear (Lush and Molln, 1942),
a linear regression seemed inadequate to correct the data properly for dif-
ferences in ages of sows. It was found by experience that squared or higher
Table 4

Classification of Litters According to Line of Sire and Year

<table>
<thead>
<tr>
<th>Year</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>S</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1942</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>9</td>
<td>6</td>
<td>11</td>
<td>0</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>47</td>
</tr>
<tr>
<td>1943</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>1944</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>1945</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>1946</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>1947</td>
<td>0</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>42</td>
</tr>
<tr>
<td>Sum</td>
<td>11</td>
<td>24</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>13</td>
<td>16</td>
<td>8</td>
<td>15</td>
<td>13</td>
<td>6</td>
<td>20</td>
<td>214</td>
</tr>
</tbody>
</table>

Table 5

Classification of Litters According to Line of Dam and Year

<table>
<thead>
<tr>
<th>Year</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>S</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1942</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>47</td>
</tr>
<tr>
<td>1943</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>14</td>
<td>41</td>
</tr>
<tr>
<td>1944</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>1945</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>1946</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>33</td>
</tr>
<tr>
<td>1947</td>
<td>0</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>42</td>
</tr>
<tr>
<td>Sum</td>
<td>15</td>
<td>21</td>
<td>27</td>
<td>28</td>
<td>13</td>
<td>8</td>
<td>12</td>
<td>14</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>55</td>
<td>214</td>
</tr>
</tbody>
</table>

Order terms make an iterative solution very difficult. Since the equations were so numerous as to require an iterative solution, any method of analysis which interfered with a rapid iterative solution was undesirable. Consequently, it was decided to classify the sows into arbitrary age groups. Since the effect of change in age was most pronounced at the younger ages and since it was desirable to keep the age classes to a minimum, the groupings chosen were first litter under 12 months, first litter more than 12
months, and second or later litter. The two-way classifications, line of sire by age of dam and line of dam by age of dam are presented in Tables 6 and 7 respectively.

Table 6

Classification of Litters According to Line of Sire and Age of Dam

<table>
<thead>
<tr>
<th>Age</th>
<th>Line of Sire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes: A</td>
<td>B</td>
</tr>
<tr>
<td>1-10 :</td>
<td>3</td>
</tr>
<tr>
<td>1-12 :</td>
<td>4</td>
</tr>
<tr>
<td>2 +  :</td>
<td>4</td>
</tr>
<tr>
<td>Sum   : 11</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 7

Classification of Litters According to Line of Dam and Age of Dam

<table>
<thead>
<tr>
<th>Age</th>
<th>Line of Dam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes: A</td>
<td>B</td>
</tr>
<tr>
<td>1-10 :</td>
<td>1</td>
</tr>
<tr>
<td>1-12 :</td>
<td>7</td>
</tr>
<tr>
<td>2 +  :</td>
<td>7</td>
</tr>
<tr>
<td>Sum   : 15</td>
<td>21</td>
</tr>
</tbody>
</table>

In addition to age classes the sows were divided into arbitrary classes with respect to their coefficients of inbreeding. The classes chosen were 15 to 25, 25 to 30, 30 to 35, 35 to 40, 40 to 50, and 50 or more. When the inbreeding coefficient fell on the division between two classes, the litter was put in the higher of the two groups. Line of sire by inbreeding of dam and line of dam by inbreeding of dam subclass numbers are presented
in Tables 8 and 9.

Table 8

Classification of Litters According to Line of Sire and Inbreeding of Dam

<table>
<thead>
<tr>
<th>Inbreeding:</th>
<th>Line of Sire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes:</td>
<td>A</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>50+</td>
<td>4</td>
</tr>
<tr>
<td>Sum</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 9

Classification of Litters According to Line of Dam and Inbreeding of Dam

<table>
<thead>
<tr>
<th>Inbreeding:</th>
<th>Line of Dam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes:</td>
<td>A</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Sum</td>
<td>15</td>
</tr>
</tbody>
</table>

Tables 2 - 9 inclusive illustrate the extreme non-orthogonality of the single cross data. A further illustration of this fact is the arrangement of subclass numbers with respect to the two-way classification, age of dam by inbreeding of dam. (Table 10)
Table 10

Classification of Litters According to Inbreeding of Dam and Age of Dam

<table>
<thead>
<tr>
<th>Inbreeding</th>
<th>Age Classes</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>1-10</td>
<td>1-12</td>
<td>2+</td>
<td>Sum</td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>15-25</td>
<td>8</td>
<td>9</td>
<td>27</td>
<td>44</td>
</tr>
<tr>
<td>25-30</td>
<td>8</td>
<td>14</td>
<td>29</td>
<td>51</td>
</tr>
<tr>
<td>30-35</td>
<td>6</td>
<td>11</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>35-40</td>
<td>11</td>
<td>6</td>
<td>18</td>
<td>35</td>
</tr>
<tr>
<td>40-50</td>
<td>5</td>
<td>12</td>
<td>16</td>
<td>33</td>
</tr>
<tr>
<td>50+</td>
<td>1</td>
<td>9</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>Sum</td>
<td>39</td>
<td>61</td>
<td>114</td>
<td>214</td>
</tr>
</tbody>
</table>

It would be expected that progeny of different boars within lines differ. If such differences do exist, apparent differences among the general combining abilities for lines are due in part to sire differences. As was described in Section II H 1, an analysis considering sire differences is much more difficult than one in which such differences are ignored. It turned out, however, in this study that there was no evidence that the progeny of different sires of the same line differed. A preliminary study with respect to each of the eight litter characteristics was made by the variance component method. The short method described in Section II F 1 was utilized to obtain estimates of the variances for line, maternal, specific, age, inbreeding, and boar effects. The estimate of the boar variance for each of the eight criteria was negative. This gave rather strong evidence that a test of significance by the least squares method would yield non-significant mean square for boars. In order to test this assumption a least squares analysis was accomplished for the measurement with the smallest negative variance estimate (21 day weight). The results of this test.
demonstrated conclusively that the mean squares for boar differences were smaller than those for error. It is not conceivable that boar differences are really non-existent. Rather, it seems plausible that the vagaries of sampling only make that appear to be the case in these data. So far as this study is concerned, failure to demonstrate boar differences merely shows that ignoring this classification in the analysis has little effect on the results.

C. Tests of Significance

The general theory of tests of significance by least squares methods and their application to analysis of single cross data were presented in Section II. Tables 11 and 12 present the computed values of the various reductions in sums of squares needed to effect the tests.

Table 11

<table>
<thead>
<tr>
<th>Reduction</th>
<th>d.f.</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>R(T)</td>
<td>214</td>
<td>10,904</td>
</tr>
<tr>
<td>R(μ, a, F, g, s, r)</td>
<td>84</td>
<td>10,199</td>
</tr>
<tr>
<td>R(μ, a, F, g, e)</td>
<td>57</td>
<td>10,121</td>
</tr>
<tr>
<td>R(μ, a, F, g, m)</td>
<td>30</td>
<td>9,888</td>
</tr>
<tr>
<td>R(μ, a, F, g)</td>
<td>19</td>
<td>9,842</td>
</tr>
<tr>
<td>R(μ, a, F)</td>
<td>8</td>
<td>9,747</td>
</tr>
</tbody>
</table>
Table 12

Reductions in Sums of Squares with Respect to Litter Weights

<table>
<thead>
<tr>
<th>Reduction</th>
<th>d.f.</th>
<th>0</th>
<th>21</th>
<th>56</th>
<th>154</th>
</tr>
</thead>
<tbody>
<tr>
<td>R(T)</td>
<td>214</td>
<td>93,199</td>
<td>781,630</td>
<td>5,810,437</td>
<td>104,121,093</td>
</tr>
<tr>
<td>R(μ, a, F, g, s, r)</td>
<td>84</td>
<td>88,779</td>
<td>700,374</td>
<td>5,201,362</td>
<td>92,563,427</td>
</tr>
<tr>
<td>R(μ, a, F, g, s)</td>
<td>57</td>
<td>87,990</td>
<td>689,076</td>
<td>5,118,032</td>
<td>91,518,308</td>
</tr>
<tr>
<td>R(μ, a, F, g, m)</td>
<td>30</td>
<td>86,187</td>
<td>667,038</td>
<td>4,945,030</td>
<td>87,581,379</td>
</tr>
<tr>
<td>R(μ, a, F, g)</td>
<td>19</td>
<td>85,871</td>
<td>657,528</td>
<td>4,897,101</td>
<td>86,537,181</td>
</tr>
<tr>
<td>R(μ, a, F)</td>
<td>8</td>
<td>84,965</td>
<td>642,725</td>
<td>4,807,612</td>
<td>85,011,335</td>
</tr>
</tbody>
</table>

The tests of hypotheses made in this study are inexact to the extent that the distribution of errors departs from normality. Two characteristics of these data make for non-normality; (1) litter number is not continuously distributed, and (2) litter number and litter weight are zero in a disproportionately large number of cases at 21, 56, and 154 days. In the case, however, of classifications as complex as these, it is hardly likely that exact tests, even if they could be made, would result in conclusions any different from those arising from the inexact tests (Cochran, 1946).

Furthermore, the estimates of line and cross line effects and the estimates of variances are unbiased no matter what may be the distribution of errors.

1. Test of hypothesis that \( \beta_{i} = \theta_{j} = 0 \)

Since the test of the hypothesis that \( m_{j} = 0 \) was accepted (Table 15), the appropriate test for the hypothesis that \( \beta_{i} = \theta_{j} = 0 \) is \( R(\mu, a, F, g, s, r) - R(\mu, a, F, g, s) \). The first of these reductions was obtained by fitting constants to age, \( F \), and reciprocal crosses after absorbing the coefficients for reciprocal crosses into those for age and \( F \) as described in Section II G 1. The second reduction was computed by fitting constants to age, \( F \), and
crosses ignoring which way the cross was made, after absorbing the coefficients of the crosses into the coefficients of the other parameters. The results of the test of the hypothesis that \( r_{ij} = 0 \) appear in Table 13.

### Table 13

Test of Hypothesis that \( r_{ij} = 0 \)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Litter Number</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d.f.</td>
<td>0</td>
</tr>
<tr>
<td>Among R</td>
<td>27</td>
<td>2.89</td>
</tr>
<tr>
<td>Error</td>
<td>130</td>
<td>5.42</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ratio (F)</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>

It is obvious that the hypothesis that \( r_{ij} = 0 \) should be accepted for each of the eight measurements, since the mean square for R is less than the error mean square in each case.

2. **Test of hypothesis that \( s_{ij} = 0 \)**

Since the hypotheses that \( r_{ij} = 0 \) and that \( m_j = 0 \) were both accepted (Tables 13 and 15), the appropriate test for the hypothesis that \( s_{ij} = 0 \) is \( R(\mu, a, F, g, s) - R(\mu, a, F, g) \). The method used for computing the first of these reductions was described in Section III C 1, and the second reduction was obtained by absorbing the coefficients of \( F \) into the coefficients of \( a \) and \( g_j \). The results of this test are presented in Table 14. The mean squares for \( S \) for litter number at birth, for litter weight at birth, and for litter weight at 154 days are all significant at the 5 percent level. Although the other mean squares are not significant at this level, they all exceed 1 enough to indicate that the specific combining ability effect is real.
Table 14

Test of Hypothesis that $s_{ij} = 0$

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>d.f.</th>
<th>Litter Number</th>
<th>Mean Squares</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>Among S</td>
<td>38</td>
<td>7.34</td>
<td>6.03</td>
<td>6.13</td>
</tr>
<tr>
<td>Error</td>
<td>157</td>
<td>4.99</td>
<td>4.97</td>
<td>4.68</td>
</tr>
<tr>
<td>Variance ratio (F)</td>
<td></td>
<td>1.47</td>
<td>1.21</td>
<td>1.31</td>
</tr>
</tbody>
</table>

$F_{0.05} = 1.47$, $F_{0.01} = 1.72$

3. Test of the hypothesis that $m_j = 0$

Strictly speaking, the appropriate test for this hypothesis is

$$R(\mu, a, F, g, m, s) - R(\mu, a, F, g, s).$$

It was found, however, that the first of these reductions was a particularly difficult one to obtain inasmuch as the iterative solution did not converge toward the correct estimates as rapidly as it does for most least squares equations. Therefore, the sum of squares, $R(\mu, a, F, g, m) - R(\mu, a, F, g)$, was utilized. The first of these reductions was computed by fitting constants to age, $F$, line of sire, and line of dam. The equations were solved after the coefficients of line of dam had been absorbed. The second reduction involved fitting constants to age, $F$, and $g_i$, the solution to the equations following absorption of the coefficients for $F$. Table 15 presents the results of this test.

Both the $M$ and error mean squares have $\sigma_s^2$ terms in their expectations. Consequently, both are biased upwards if it is assumed that $\sigma_s^2 > 0$. Even with this bias present in the mean square for $M$, the true mean square for error cannot possibly be small enough to make among $M$ significant. Note,
for example, the sizes of the mean squares for error in Tables 13 and 14. Therefore, the hypothesis that \( m_j = 0 \) was accepted for each of the litter measurements with the possible exception of 21 day litter weight.

Table 15
Test of Hypothesis that \( m_j = 0 \)

<table>
<thead>
<tr>
<th>Source of Variation: d.f.</th>
<th>Litter Number</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21 56 154</td>
<td>0 21 56 154</td>
</tr>
<tr>
<td>Among M</td>
<td>11 4.18 4.45 3.91 4.36</td>
<td>29 865 4358 94,927</td>
</tr>
<tr>
<td>Error</td>
<td>184 5.52 5.22 5.03 4.69</td>
<td>38 623 4703 89,890</td>
</tr>
<tr>
<td>Variance ratio (F)</td>
<td>&lt; 1 &lt; 1 &lt; 1 &lt; 1 1.39 &lt; 1 1.06</td>
<td></td>
</tr>
<tr>
<td>( F.05 = 1.84 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Test of the hypothesis that \( g_i = 0 \)

Assuming that \( r_{ij} = m_j = 0 \), and that \( s_{ij} \neq 0 \), the strictly correct test for the hypothesis that \( g_i = 0 \) is \( R(\mu, a, F, g, s) - R(\mu, a, F, g) \). This test is, however, quite impractical since the minimum number of equations requiring solution would be 40. Consequently, the test actually used was \( R(\mu, a, F, g) - R(\mu, a, F) \). The results of this analysis are presented in Table 16.

Table 16
Test of Hypothesis that \( g_i = 0 \)

<table>
<thead>
<tr>
<th>Source of Variation: d.f.</th>
<th>Litter Number</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21 56 154</td>
<td>0 21 56 154</td>
</tr>
<tr>
<td>Among G</td>
<td>11 8.64 10.00 9.00 7.64</td>
<td>82 1346 8117 138,713</td>
</tr>
<tr>
<td>Error</td>
<td>195 5.45 5.18 4.96 4.67</td>
<td>38 636 4684 90,174</td>
</tr>
<tr>
<td>Variance ratio (F)</td>
<td>1.59 1.93 1.81 1.64 2.16 2.12 1.73 1.53</td>
<td></td>
</tr>
<tr>
<td>( F.05 = 1.83, F.01 = 2.34 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The \( F \) values are large enough in most cases to indicate that the differences in general combining abilities among the lines are real. It should be recognized, however, that the expectation of the mean square for \( G \) contains \( \sum s^2 \), as does the error mean square. If the mean square for \( S \) is used as the error term, the \( F \) values for the eight different measurements are 1.18, 1.66, 1.47, 1.30, 1.46, 1.62, 1.40, and 1.06, the \( F \) value required for significance at the 5 per cent level being 2.05. Consequently, none of the mean squares is significant when this test is made, but there is still a suggestion that general combining abilities differ among lines. Additional information on general combining abilities will be presented in connection with estimates of variances.

D. Least Squares Estimates

Since the hypotheses that \( r_{ij} = 0 \) and that \( m_j = 0 \) were both accepted and since there was rather strong evidence that \( s_{ij} \neq 0 \), the appropriate analysis from which to obtain estimates of \( g_i \) and of \( g_i + g_j + s_{ij} \) is \( R(\mu, a, F, g, s) \). Estimates of \( g_i \) with considerably smaller sampling errors arise in \( R(\mu, a, F, g) \), but the estimates are biased if it is assumed that \( s_{ij} \neq 0 \).

The estimates of \( \mu + g_i + g_j + s_{ij} \) for each of the crosses tested are given in Table 17. These estimates are, of course, subject to very large sampling errors inasmuch as the number of litters per cross was small. These few negative estimates have no biological meaning, but are merely a consequence of only one litter of that particular cross being tested, all of the pigs having died before the age of 21 days, and the sum of correction factors for age and inbreeding being less than 0.
<table>
<thead>
<tr>
<th>Cross</th>
<th>Litter Number</th>
<th>Litter Weight (lbs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>AB</td>
<td>9.6</td>
<td>6.7</td>
</tr>
<tr>
<td>AC</td>
<td>4.3</td>
<td>4.2</td>
</tr>
<tr>
<td>AD</td>
<td>6.1</td>
<td>5.7</td>
</tr>
<tr>
<td>AE</td>
<td>6.6</td>
<td>7.2</td>
</tr>
<tr>
<td>AF</td>
<td>7.8</td>
<td>5.6</td>
</tr>
<tr>
<td>AH</td>
<td>5.4</td>
<td>4.2</td>
</tr>
<tr>
<td>AI</td>
<td>6.5</td>
<td>4.9</td>
</tr>
<tr>
<td>AJ</td>
<td>7.1</td>
<td>5.7</td>
</tr>
<tr>
<td>AS</td>
<td>8.0</td>
<td>5.2</td>
</tr>
<tr>
<td>BG</td>
<td>6.2</td>
<td>3.7</td>
</tr>
<tr>
<td>BD</td>
<td>3.5</td>
<td>3.2</td>
</tr>
<tr>
<td>BE</td>
<td>6.9</td>
<td>6.3</td>
</tr>
<tr>
<td>BG</td>
<td>5.6</td>
<td>3.5</td>
</tr>
<tr>
<td>BH</td>
<td>2.6</td>
<td>1.8</td>
</tr>
<tr>
<td>BI</td>
<td>2.7</td>
<td>-0.6</td>
</tr>
<tr>
<td>BJ</td>
<td>5.2</td>
<td>4.2</td>
</tr>
<tr>
<td>BK</td>
<td>7.7</td>
<td>3.4</td>
</tr>
<tr>
<td>BS</td>
<td>6.4</td>
<td>4.3</td>
</tr>
<tr>
<td>CD</td>
<td>6.2</td>
<td>5.3</td>
</tr>
<tr>
<td>GE</td>
<td>6.9</td>
<td>6.4</td>
</tr>
<tr>
<td>CF</td>
<td>5.6</td>
<td>6.2</td>
</tr>
<tr>
<td>CG</td>
<td>8.0</td>
<td>3.2</td>
</tr>
<tr>
<td>CH</td>
<td>6.6</td>
<td>5.4</td>
</tr>
<tr>
<td>CJ</td>
<td>6.4</td>
<td>5.3</td>
</tr>
<tr>
<td>CK</td>
<td>6.2</td>
<td>4.5</td>
</tr>
<tr>
<td>CS</td>
<td>5.9</td>
<td>4.8</td>
</tr>
<tr>
<td>DE</td>
<td>6.4</td>
<td>4.9</td>
</tr>
<tr>
<td>DH</td>
<td>2.8</td>
<td>3.6</td>
</tr>
<tr>
<td>DI</td>
<td>7.2</td>
<td>6.0</td>
</tr>
<tr>
<td>DJ</td>
<td>6.3</td>
<td>2.4</td>
</tr>
<tr>
<td>DK</td>
<td>4.0</td>
<td>1.1</td>
</tr>
<tr>
<td>DS</td>
<td>5.8</td>
<td>4.5</td>
</tr>
<tr>
<td>EF</td>
<td>5.8</td>
<td>4.7</td>
</tr>
<tr>
<td>EG</td>
<td>6.0</td>
<td>4.2</td>
</tr>
<tr>
<td>EI</td>
<td>5.4</td>
<td>5.6</td>
</tr>
<tr>
<td>ES</td>
<td>6.2</td>
<td>4.8</td>
</tr>
<tr>
<td>FH</td>
<td>6.6</td>
<td>3.1</td>
</tr>
<tr>
<td>FI</td>
<td>5.2</td>
<td>4.4</td>
</tr>
<tr>
<td>FJ</td>
<td>5.1</td>
<td>4.2</td>
</tr>
<tr>
<td>FS</td>
<td>7.9</td>
<td>7.2</td>
</tr>
<tr>
<td>GH</td>
<td>7.4</td>
<td>2.9</td>
</tr>
<tr>
<td>GI</td>
<td>8.2</td>
<td>5.2</td>
</tr>
</tbody>
</table>
Table 17 (Cont'd.)

<table>
<thead>
<tr>
<th>Cross</th>
<th>Litter Number</th>
<th>Litter Weight (lbs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>GJ</td>
<td>9.3</td>
<td>7.5</td>
</tr>
<tr>
<td>GK</td>
<td>4.6</td>
<td>-2</td>
</tr>
<tr>
<td>GS</td>
<td>6.7</td>
<td>4.6</td>
</tr>
<tr>
<td>HS</td>
<td>5.6</td>
<td>6.2</td>
</tr>
<tr>
<td>IS</td>
<td>6.4</td>
<td>4.8</td>
</tr>
<tr>
<td>JS</td>
<td>7.6</td>
<td>6.2</td>
</tr>
<tr>
<td>IK</td>
<td>3.7</td>
<td>4.6</td>
</tr>
<tr>
<td>KS</td>
<td>6.3</td>
<td>6.2</td>
</tr>
<tr>
<td>(\hat{\mu})</td>
<td>6.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Of more practical value than the above estimates are the estimates of \(g_i\), which arise from the \(\mu + g_i + g_j + a_{ij}\). These estimates were affected in the manner described in Section 5.2.c and are presented in Table 18.

Table 18

<table>
<thead>
<tr>
<th>Line</th>
<th>Litter Number</th>
<th>Litter Weight (lbs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>A</td>
<td>7.0</td>
<td>5.4</td>
</tr>
<tr>
<td>B</td>
<td>5.6</td>
<td>3.5</td>
</tr>
<tr>
<td>C</td>
<td>6.1</td>
<td>4.9</td>
</tr>
<tr>
<td>D</td>
<td>5.5</td>
<td>4.0</td>
</tr>
<tr>
<td>E</td>
<td>6.0</td>
<td>5.4</td>
</tr>
<tr>
<td>F</td>
<td>6.2</td>
<td>4.6</td>
</tr>
<tr>
<td>G</td>
<td>7.2</td>
<td>3.9</td>
</tr>
<tr>
<td>H</td>
<td>5.0</td>
<td>3.8</td>
</tr>
<tr>
<td>I</td>
<td>5.5</td>
<td>4.4</td>
</tr>
<tr>
<td>J</td>
<td>6.4</td>
<td>5.0</td>
</tr>
<tr>
<td>K</td>
<td>5.4</td>
<td>3.5</td>
</tr>
<tr>
<td>S</td>
<td>6.6</td>
<td>5.4</td>
</tr>
<tr>
<td>(\hat{\mu})</td>
<td>6.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>
Although the litter was considered the experimental unit in this study, it is possible to compute estimates of individual pig weights from the litter estimates. For example, estimates of $g_1$ with respect to individual weight can be obtained by dividing $\hat{\mu} + \hat{\delta}_1$ for weight at a given age by $\hat{\mu} + \hat{\delta}_1$ for number at the same age. Table 19 presents the results of such an analysis. Estimates with smaller sampling errors could have been obtained by setting up separate least squares equations for each of the ages, using the pig as the experimental unit rather than the litter. However, as was previously mentioned, such an analysis would require a separate set of equations for each age. The estimates of Table 19 appear adequate for the present purposes.

Table 19

<table>
<thead>
<tr>
<th>Line</th>
<th>Age</th>
<th>(weight in pounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>A</td>
<td>2.7</td>
<td>9.5</td>
</tr>
<tr>
<td>B</td>
<td>2.8</td>
<td>10.6</td>
</tr>
<tr>
<td>C</td>
<td>2.8</td>
<td>10.9</td>
</tr>
<tr>
<td>D</td>
<td>2.6</td>
<td>10.2</td>
</tr>
<tr>
<td>E</td>
<td>2.9</td>
<td>10.8</td>
</tr>
<tr>
<td>F</td>
<td>2.6</td>
<td>9.9</td>
</tr>
<tr>
<td>G</td>
<td>2.6</td>
<td>9.7</td>
</tr>
<tr>
<td>H</td>
<td>3.0</td>
<td>11.7</td>
</tr>
<tr>
<td>I</td>
<td>2.8</td>
<td>11.9</td>
</tr>
<tr>
<td>J</td>
<td>3.2</td>
<td>11.5</td>
</tr>
<tr>
<td>K</td>
<td>2.9</td>
<td>11.6</td>
</tr>
<tr>
<td>S</td>
<td>2.9</td>
<td>11.5</td>
</tr>
</tbody>
</table>

It is also possible by utilizing the least squares estimates of Table 18 to obtain estimates of livability for each of the lines. These estimates
are subject to larger sampling errors than estimates obtained by solving least squares equations using the pig as the experimental unit, but the additional labor required in the latter procedure is scarcely justified in this study. Table 20 presents the fraction of pigs which survived between certain ages. For example, in the case of Line A the survival from birth to 21 days is obtained by dividing 5.4, the number of pigs per litter alive at 21 days, by 7.0, the number of pigs per litter at birth. The entries of Table 20 were actually obtained from estimates which had not been rounded as much as the entries in Table 18.

Table 20

<table>
<thead>
<tr>
<th>Age Interval</th>
<th>Proportion surviving</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Line</td>
<td>0-21</td>
</tr>
<tr>
<td>A</td>
<td>0.77</td>
</tr>
<tr>
<td>B</td>
<td>0.63</td>
</tr>
<tr>
<td>C</td>
<td>0.80</td>
</tr>
<tr>
<td>D</td>
<td>0.73</td>
</tr>
<tr>
<td>E</td>
<td>0.90</td>
</tr>
<tr>
<td>F</td>
<td>0.75</td>
</tr>
<tr>
<td>G</td>
<td>0.54</td>
</tr>
<tr>
<td>H</td>
<td>0.77</td>
</tr>
<tr>
<td>I</td>
<td>0.80</td>
</tr>
<tr>
<td>J</td>
<td>0.78</td>
</tr>
<tr>
<td>K</td>
<td>0.64</td>
</tr>
<tr>
<td>S</td>
<td>0.81</td>
</tr>
</tbody>
</table>

As an aid in appraising the different lines Table 21 was prepared in which lines were ranked with respect to some of the more important characteristics, namely, number at birth, survival from birth to 154 days, number at 154 days, individual pig weight at 154 days, and litter weight at 154 days. The appraisal of line differences is dependent also upon the amount
by which each of the least squares estimates is regressed toward 0.

Table 21

<table>
<thead>
<tr>
<th>Ranks at birth</th>
<th>Number of pigs at 0-154</th>
<th>Number of pigs at 154</th>
<th>Individual pig weight at 154</th>
<th>Litter weight at 154</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>E</td>
<td>E</td>
<td>G</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>C</td>
<td>S</td>
<td>J</td>
</tr>
<tr>
<td>3</td>
<td>S</td>
<td>I</td>
<td>J</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>J</td>
<td>S</td>
<td>G</td>
<td>I</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>J</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>H</td>
<td>I</td>
<td>H</td>
</tr>
<tr>
<td>7</td>
<td>E</td>
<td>A</td>
<td>F</td>
<td>S</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>9</td>
<td>I</td>
<td>F</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>D</td>
<td>K</td>
<td>B</td>
<td>F</td>
</tr>
<tr>
<td>11</td>
<td>K</td>
<td>B</td>
<td>K</td>
<td>B</td>
</tr>
<tr>
<td>12</td>
<td>H</td>
<td>G</td>
<td>H</td>
<td>K</td>
</tr>
</tbody>
</table>

Although this study was not designed specifically to study the effects of age of the sow and inbreeding of the sow on litter and individual pig characteristics, the estimation of such effects was an automatic consequence of the least squares analysis. In Table 22 are presented the means for the various age classes and inbreeding classes, the former having been corrected for F, G, and S, and the latter for a, g, and s. In Table 23 individual pig weights for each of the age and inbreeding classes are given. Table 24 presents estimates of survival during various periods for different age classes and inbreeding classes. These statistics were computed in the same manner as were the individual weight and survival estimates for lines.
## Table 22

Least Squares Estimates of Age and Inbreeding Effects on Litter Number and Weight

<table>
<thead>
<tr>
<th>Litter Number</th>
<th>Litter Weight (lbs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\alpha} ) + ( \hat{\beta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st litter, under 1 yr.</td>
<td>4.9</td>
</tr>
<tr>
<td>1st litter, over 1 yr.</td>
<td>6.3</td>
</tr>
<tr>
<td>2nd or later litter</td>
<td>6.9</td>
</tr>
</tbody>
</table>

\( \hat{\alpha} \)

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\alpha} ) + ( \hat{\beta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 - 25</td>
<td>6.9</td>
</tr>
<tr>
<td>25 - 30</td>
<td>6.3</td>
</tr>
<tr>
<td>30 - 35</td>
<td>5.9</td>
</tr>
<tr>
<td>35 - 40</td>
<td>6.8</td>
</tr>
<tr>
<td>40 - 50</td>
<td>5.9</td>
</tr>
<tr>
<td>50+</td>
<td>4.5</td>
</tr>
</tbody>
</table>

## Table 23

Estimates of Age and Inbreeding Effects on Individual Pig Weights

<table>
<thead>
<tr>
<th>Age classes</th>
<th>Age</th>
<th>( \hat{\alpha} ) + ( \hat{\beta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st litter, under 1 yr.</td>
<td>2.7</td>
<td>9.8</td>
</tr>
<tr>
<td>1st litter, over 1 yr.</td>
<td>2.7</td>
<td>10.2</td>
</tr>
<tr>
<td>2nd or later litter</td>
<td>3.0</td>
<td>12.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inbreeding dams</th>
<th>( \hat{\alpha} ) + ( \hat{\beta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 - 25</td>
<td>3.0</td>
</tr>
<tr>
<td>25 - 30</td>
<td>3.0</td>
</tr>
<tr>
<td>30 - 35</td>
<td>2.9</td>
</tr>
<tr>
<td>35 - 40</td>
<td>2.7</td>
</tr>
<tr>
<td>40 - 50</td>
<td>2.7</td>
</tr>
<tr>
<td>50+</td>
<td>2.4</td>
</tr>
</tbody>
</table>
Table 24

Estimates of Age and Inbreeding Effects on Survival

<table>
<thead>
<tr>
<th>Age Class</th>
<th>Age Interval</th>
<th>0-21</th>
<th>0-56</th>
<th>0-154</th>
<th>21-56</th>
<th>21-154</th>
<th>56-154</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st litter under 1 yr.</td>
<td>15 - 25</td>
<td>.77</td>
<td>.75</td>
<td>.67</td>
<td>.96</td>
<td>.86</td>
<td>.90</td>
</tr>
<tr>
<td>1st litter over 1 yr.</td>
<td>25 - 30</td>
<td>.73</td>
<td>.69</td>
<td>.62</td>
<td>.95</td>
<td>.84</td>
<td>.89</td>
</tr>
<tr>
<td>2nd or later litter</td>
<td>35 - 40</td>
<td>.73</td>
<td>.72</td>
<td>.70</td>
<td>.99</td>
<td>.96</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>40 - 50</td>
<td>.72</td>
<td>.71</td>
<td>.66</td>
<td>.98</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>50+</td>
<td>.71</td>
<td>.70</td>
<td>.68</td>
<td>.76</td>
<td>.84</td>
<td>.88</td>
</tr>
</tbody>
</table>

E. Sampling Errors of the Least Squares Estimates

Under the assumption that $m_j = r_{ij} = 0$ sampling errors were calculated for the least squares estimates presented in Section III D. These computations were done in the manner described in Section II G 3 c. The errors are expressed as the square root of the variances of the various parameter estimates. These statistics are presented in Tables 25, 26, and 27.

Table 25

Standard Deviations of $\hat{A} + \hat{e}_i + \hat{g}_j + \hat{e}_{ij}$, Assuming $m_j = r_{ij} = 0$

<table>
<thead>
<tr>
<th>Gross</th>
<th>Litter Number</th>
<th>Litter Weight (lbs.)</th>
<th>0 &amp; 21</th>
<th>56</th>
<th>154</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>21</td>
<td>56</td>
</tr>
<tr>
<td>AB</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.2</td>
<td>18</td>
</tr>
<tr>
<td>AC</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>AD</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.2</td>
<td>18</td>
</tr>
<tr>
<td>AE</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>AF</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.7</td>
<td>11</td>
</tr>
<tr>
<td>AH</td>
<td>1.3</td>
<td>1.3</td>
<td>1.2</td>
<td>3.4</td>
<td>14</td>
</tr>
<tr>
<td>AI</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>3.0</td>
<td>13</td>
</tr>
<tr>
<td>Cross</td>
<td>Litter Number</td>
<td>Litter Weight (lbs.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 &amp; 21</td>
<td>56</td>
<td>154</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>AJ</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>3.5</td>
<td>15</td>
</tr>
<tr>
<td>AS</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.6</td>
<td>11</td>
</tr>
<tr>
<td>BC</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.6</td>
<td>11</td>
</tr>
<tr>
<td>BD</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>2.4</td>
<td>10</td>
</tr>
<tr>
<td>BE</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>2.4</td>
<td>10</td>
</tr>
<tr>
<td>BG</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>3.0</td>
<td>13</td>
</tr>
<tr>
<td>BH</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.2</td>
<td>18</td>
</tr>
<tr>
<td>BI</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>BJ</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>BK</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>BS</td>
<td>.6</td>
<td>.5</td>
<td>.5</td>
<td>1.4</td>
<td>6</td>
</tr>
<tr>
<td>CD</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>1.9</td>
<td>8</td>
</tr>
<tr>
<td>CK</td>
<td>.8</td>
<td>.8</td>
<td>.8</td>
<td>2.1</td>
<td>9</td>
</tr>
<tr>
<td>CF</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>GG</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.7</td>
<td>11</td>
</tr>
<tr>
<td>GH</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>2.5</td>
<td>10</td>
</tr>
<tr>
<td>GJ</td>
<td>1.3</td>
<td>1.3</td>
<td>1.2</td>
<td>3.4</td>
<td>14</td>
</tr>
<tr>
<td>GK</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
<td>3.0</td>
<td>13</td>
</tr>
<tr>
<td>GS</td>
<td>.7</td>
<td>.7</td>
<td>.6</td>
<td>1.8</td>
<td>8</td>
</tr>
<tr>
<td>DK</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>1.8</td>
<td>8</td>
</tr>
<tr>
<td>DH</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>DI</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.2</td>
<td>18</td>
</tr>
<tr>
<td>DJ</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>2.4</td>
<td>10</td>
</tr>
<tr>
<td>DK</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.2</td>
<td>18</td>
</tr>
<tr>
<td>DS</td>
<td>.6</td>
<td>.6</td>
<td>.5</td>
<td>1.5</td>
<td>6</td>
</tr>
<tr>
<td>EF</td>
<td>1.7</td>
<td>1.6</td>
<td>1.6</td>
<td>4.3</td>
<td>18</td>
</tr>
<tr>
<td>EG</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>3.0</td>
<td>12</td>
</tr>
<tr>
<td>EI</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>ES</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>1.9</td>
<td>8</td>
</tr>
<tr>
<td>FH</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.7</td>
<td>11</td>
</tr>
<tr>
<td>FI</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
<td>3.1</td>
<td>13</td>
</tr>
<tr>
<td>FJ</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.1</td>
<td>17</td>
</tr>
<tr>
<td>FS</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.1</td>
<td>17</td>
</tr>
<tr>
<td>GH</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>3.0</td>
<td>12</td>
</tr>
<tr>
<td>GI</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.6</td>
<td>11</td>
</tr>
<tr>
<td>GJ</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.2</td>
<td>18</td>
</tr>
<tr>
<td>GJ</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.2</td>
<td>18</td>
</tr>
<tr>
<td>GS</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>4.2</td>
<td>17</td>
</tr>
<tr>
<td>HS</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>IS</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>2.4</td>
<td>10</td>
</tr>
<tr>
<td>JS</td>
<td>1.4</td>
<td>1.3</td>
<td>1.3</td>
<td>3.5</td>
<td>15</td>
</tr>
<tr>
<td>IK</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
<tr>
<td>KS</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>5.9</td>
<td>25</td>
</tr>
</tbody>
</table>
Table 26

Standard Deviations of \( \hat{\sigma}_1 \), Assuming \( m_j = r_{ij} = 0 \)

<table>
<thead>
<tr>
<th>Line</th>
<th>Litter Number</th>
<th>Litter Weight (lbs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 &amp; 21</td>
<td>56</td>
</tr>
<tr>
<td>A</td>
<td>.56</td>
<td>.54</td>
</tr>
<tr>
<td>B</td>
<td>.52</td>
<td>.50</td>
</tr>
<tr>
<td>C</td>
<td>.47</td>
<td>.46</td>
</tr>
<tr>
<td>D</td>
<td>.49</td>
<td>.48</td>
</tr>
<tr>
<td>E</td>
<td>.48</td>
<td>.47</td>
</tr>
<tr>
<td>F</td>
<td>.62</td>
<td>.60</td>
</tr>
<tr>
<td>G</td>
<td>.53</td>
<td>.51</td>
</tr>
<tr>
<td>H</td>
<td>.62</td>
<td>.60</td>
</tr>
<tr>
<td>I</td>
<td>.62</td>
<td>.60</td>
</tr>
<tr>
<td>J</td>
<td>.60</td>
<td>.58</td>
</tr>
<tr>
<td>K</td>
<td>.77</td>
<td>.75</td>
</tr>
<tr>
<td>S</td>
<td>.47</td>
<td>.45</td>
</tr>
</tbody>
</table>

Table 27

Standard Deviations of \( \hat{\alpha} \) and \( \hat{\sigma} \), Assuming \( m_j = r_{ij} = 0 \)

<table>
<thead>
<tr>
<th>Age Classes</th>
<th>Litter Number</th>
<th>Litter Weight (lbs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 &amp; 21</td>
<td>56</td>
</tr>
<tr>
<td>1st litter, under 1 yr.</td>
<td>.31</td>
<td>.30</td>
</tr>
<tr>
<td>1st litter, over 1 yr.</td>
<td>.28</td>
<td>.27</td>
</tr>
<tr>
<td>2nd or later litter</td>
<td>.24</td>
<td>.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inbreeding Classes</th>
<th>Litter Number</th>
<th>Litter Weight (lbs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 &amp; 21</td>
<td>56</td>
</tr>
<tr>
<td>15 - 25</td>
<td>.43</td>
<td>.42</td>
</tr>
<tr>
<td>25 - 30</td>
<td>.36</td>
<td>.35</td>
</tr>
<tr>
<td>30 - 35</td>
<td>.39</td>
<td>.38</td>
</tr>
<tr>
<td>35 - 40</td>
<td>.41</td>
<td>.40</td>
</tr>
<tr>
<td>40 - 50</td>
<td>.41</td>
<td>.40</td>
</tr>
<tr>
<td>50+</td>
<td>.58</td>
<td>.57</td>
</tr>
</tbody>
</table>

F. Estimates of Variances

Early in the study the only estimates of the variances which appeared obtainable were those arising from an analysis of variance ignoring the
non-orthogonality. The labor of inverting the matrices to obtain expected
values of least squares sums of squares seemed too great to justify using
that method of estimating variances. Later, however, the computational
short cuts described in Section II < 4 were developed, and these made it
possible to estimate the variances by the least squares procedure. The es-
timates computed by ignoring the non-orthogonality are presented here merely
to illustrate the method and to demonstrate the fact that different esti-
mates of variances are obtained depending upon what method is used. There
is, in other words, no unique solution for the estimate of a variance.

Table 28 shows the expected values of certain sums of squares computed
by ignoring the non-orthogonality. By equating these expected values to
the numerical values of the sums of squares and solving the resulting equa-
tions the estimates of variances presented in Table 29 were obtained. These
estimates are probably very poor ones for several different reasons. First,
it is assumed that the expected values of all cross products are equal to
zero, but this is most certainly not true with respect to the cross products
among the age groups and among the inbreeding groups. Second, the various
sums of squares each have in their expected values a sizeable proportion of
most of the variances other than the one which is to be obtained from that
particular sum of squares. Third, the estimates themselves do not coincide
in a logical manner with the mean squares computed by the least squares
method. This is particularly true with respect to the estimates of $\sigma_m^2$. For example, positive values were obtained for the estimates of six of the
eight categories, whereas the tests of hypotheses indicated that the only
measurement in which differences with respect to maternal effect existed
was 21 day weight.
Table 28

Expected Values of Certain Sums of Squares Calculated by Ignoring Non-Orthogonality (coefficients rounded to nearest whole number)

<table>
<thead>
<tr>
<th>Sums of Squares</th>
<th>( g )</th>
<th>( m )</th>
<th>( s )</th>
<th>( a )</th>
<th>( f )</th>
<th>( e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line of Sire</td>
<td>171</td>
<td>20</td>
<td>40</td>
<td>9</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Line of Dam</td>
<td>165</td>
<td>186</td>
<td>34</td>
<td>13</td>
<td>37</td>
<td>11</td>
</tr>
<tr>
<td>Crosses</td>
<td>338</td>
<td>115</td>
<td>206</td>
<td>28</td>
<td>51</td>
<td>49</td>
</tr>
<tr>
<td>Age Groups</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>129</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>F Groups</td>
<td>20</td>
<td>19</td>
<td>7</td>
<td>4</td>
<td>175</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>338</td>
<td>186</td>
<td>206</td>
<td>129</td>
<td>175</td>
<td>213</td>
</tr>
</tbody>
</table>

Table 29

Estimates of Variances from Expected Values of Sums of Squares Computed by Ignoring Non-Orthogonality

<table>
<thead>
<tr>
<th>Variance</th>
<th>Litter Number</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>( \sigma_g^2 )</td>
<td>-.26</td>
<td>.04</td>
</tr>
<tr>
<td>( \sigma_m^2 )</td>
<td>.54</td>
<td>.14</td>
</tr>
<tr>
<td>( \sigma_s^2 )</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>( \sigma_a^2 )</td>
<td>.55</td>
<td>.35</td>
</tr>
<tr>
<td>( \sigma_f^2 )</td>
<td>.27</td>
<td>.14</td>
</tr>
<tr>
<td>( \sigma_e^2 )</td>
<td>5.35</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Table 30

Variance Estimates of Table 29 Expressed as Fractions of \( \sigma_e^2 \)

<table>
<thead>
<tr>
<th>Variance</th>
<th>Litter Number</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>( \delta_g^2 )</td>
<td>-.05</td>
<td>.01</td>
</tr>
<tr>
<td>( \delta_m^2 )</td>
<td>.10</td>
<td>.03</td>
</tr>
<tr>
<td>( \delta_s^2 )</td>
<td>.05</td>
<td>.05</td>
</tr>
</tbody>
</table>
Although the method of least squares probably provides better estimates of the variances than the method just described, even the former offers no unique method of estimation, there being available a number of different sums of squares which might be used for estimating the variances. Just which sums of squares to choose is a matter of judgement and of ease in computation. Unbiased estimates are obtained no matter which ones are chosen, but there is a single best choice from the standpoint of efficiency. There is, however, no way of knowing the relative efficiency of the different choices. The tests of hypotheses offer some guidance with respect to this problem; for example, the fact that \( r_{ij} = 0 \) and that \( m_j = 0 \) with the possible exception of 21 day weight, suggests that the best estimates of \( \sigma_g^2, \sigma_s^2 \), and \( \sigma_e^2 \) arise from sums of squares calculated under the assumption that \( r_{ij} = m_j = 0 \). Table 31 shows the expected values of certain sums of squares computed by the method of least squares.

### Table 31

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>( \Delta g )</th>
<th>( \Delta m )</th>
<th>( \Delta s )</th>
<th>( \Delta e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R(\mu, a, F, g, m) - R(\mu, a, F, m) )</td>
<td>162</td>
<td>0</td>
<td>41</td>
<td>11</td>
</tr>
<tr>
<td>( R(\mu, a, F, g) - R(\mu, a, F) )</td>
<td>312</td>
<td>81</td>
<td>59</td>
<td>11</td>
</tr>
<tr>
<td>( R(\mu, a, F, g, m) - R(\mu, a, F, g) )</td>
<td>0</td>
<td>84</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>( R(\mu, a, F, g, s) - R(\mu, a, F, g) )</td>
<td>0</td>
<td>22</td>
<td>138</td>
<td>38</td>
</tr>
<tr>
<td>( R(T) - R(\mu, a, F, g, m) )</td>
<td>0</td>
<td>0</td>
<td>126</td>
<td>184</td>
</tr>
<tr>
<td>( R(T) - R(\mu, a, F, g, s) )</td>
<td>0</td>
<td>62</td>
<td>0</td>
<td>157</td>
</tr>
</tbody>
</table>
The variance estimates of Table 32 were obtained under the assumption that $\sigma_m^2 \neq 0$. These estimates were made in order to check whether the results thereby obtained would correspond to the tests of hypotheses. It will be noted that the estimates of $\sigma_m^2$ were negative for all measurements except 21 day weight. This corresponds perfectly to the tests of hypotheses in which mean squares for $M$ were less than error mean square on all measurements except 21 day weight. The sums of squares used in obtaining the estimates of Table 32 were as follows:

1. $R(\mu, a, F, g, m) - R(\mu, a, F, m)$
2. $R(\mu, a, F, g, m) - R(\mu, a, F, g)$
3. $R(\mu, a, F, g, s) - R(\mu, a, F, g)$
4. $R(T) - R(\mu, a, F, g, s)$.

Table 32

Estimates of Variances from Least Squares Analyses,
Assuming $\sigma_m^2 \neq 0$; $\sigma_e^2 = 0$

<table>
<thead>
<tr>
<th>Variance</th>
<th>Litter Number</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>$\sigma_g^2$</td>
<td>-24</td>
<td>.14</td>
</tr>
<tr>
<td>$\sigma_m^2$</td>
<td>-23</td>
<td>-.12</td>
</tr>
<tr>
<td>$\sigma_s^2$</td>
<td>.65</td>
<td>.30</td>
</tr>
<tr>
<td>$\sigma_e^2$</td>
<td>5.08</td>
<td>5.02</td>
</tr>
</tbody>
</table>

Table 33 presents the estimates of Table 32 as fractions of $\sigma_e^2$. 
Table 33
Variance Estimates of Table 32 Expressed as Fractions of $\sigma_e^2$

<table>
<thead>
<tr>
<th>Variance Estimated</th>
<th>Litter Number</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>$\sigma_g^2$</td>
<td>-.05</td>
<td>.03</td>
</tr>
<tr>
<td>$\sigma_m^2$</td>
<td>-.05</td>
<td>-.02</td>
</tr>
<tr>
<td>$\sigma_s^2$</td>
<td>.13</td>
<td>.06</td>
</tr>
</tbody>
</table>

The estimates of Table 32 are probably the "best" only with respect to 21 day weight. Probably better estimates for the other measurements are obtained by assuming that $\sigma_m^2 = 0$. Under this assumption the estimates presented in Table 34 were obtained. The sums of squares utilized in this case were

1. $R(\mu, a, F, g) - R(\mu, a, F)$
2. $R(\mu, a, F, g, s) - R(\mu, a, F, g)$
3. $R(T) - R(\mu, a, F, g, s)$

Table 34
Estimates of Variances from Least Squares Analysis
Assuming $\sigma_m^2 = \sigma_r^2 = 0$

<table>
<thead>
<tr>
<th>Variance Estimated</th>
<th>Litter Number</th>
<th>Litter Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>$\sigma_g^2$</td>
<td>.01</td>
<td>.12</td>
</tr>
<tr>
<td>$\sigma_s^2$</td>
<td>.64</td>
<td>.29</td>
</tr>
<tr>
<td>$\sigma_s^2$</td>
<td>4.99</td>
<td>4.97</td>
</tr>
</tbody>
</table>
The estimates of Table 34 are expressed in Table 35 as fractions of \( \sigma_e^2 \).

### Table 35

Variance Estimates of Table 34 Expressed as Fractions of \( \sigma_e^2 \)

<table>
<thead>
<tr>
<th>Variance</th>
<th>Litter Number</th>
<th></th>
<th></th>
<th></th>
<th>Litter Weight</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated</td>
<td>0</td>
<td>21</td>
<td>56</td>
<td>154</td>
<td>0</td>
<td>21</td>
<td>56</td>
<td>154</td>
</tr>
<tr>
<td>( \sigma_g^2 )</td>
<td>0</td>
<td>.02</td>
<td>.02</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.01</td>
<td>-.01</td>
</tr>
<tr>
<td>( \sigma_s^2 )</td>
<td>.13</td>
<td>.06</td>
<td>.09</td>
<td>.09</td>
<td>.19</td>
<td>.11</td>
<td>.09</td>
<td>.17</td>
</tr>
</tbody>
</table>

### G. Adjustment of Least Squares Estimates

The least squares estimates of \( g_1 \) have been adjusted toward zero in accordance with the method described in Section II G 4. In making these adjustments it was assumed that \( m_j = r_{ij} = 0 \). The appropriate regression is:

\[
b(g_1 \hat{g}_1) = \frac{p-1}{p} \sigma_e^2 + \sum_{ij} \frac{1}{g_{ij}} \sigma_s^2 + \frac{1}{g_e} \sigma_e^2.
\]

The variance estimates used for calculating these regressions are those which appear in Table 34, the estimates assuming \( \sigma_m^2 = \sigma_r^2 = 0 \). \( \frac{p-1}{p} \) is, of course, \( \frac{11}{12} \). The \( g_1^1 \) are obtained from the inverse matrix for \( R(\mu, a, F, g, s) \), and their values are:

\[
A = .062 \quad E = .047 \quad I = .076
\]
\[
B = .054 \quad F = .077 \quad J = .071
\]
\[
C = .045 \quad G = .056 \quad K = .120
\]
\[
D = .049 \quad H = .078 \quad S = .044.
\]
The coefficients of \( \delta_s^2 \) in the denominator are extremely tedious to compute, consequently, an approximation was used. The error in so doing is very slight, for it will be noted that the approximate coefficient multiplied by \( \delta_s^2 \) is very small compared to \( g_1 \delta_s^2 \). The approximation used was the value of the coefficient of \( \delta_s^2 \) in a balanced design, namely,

\[
\frac{(p-1) \delta_s^2}{p(p-2)} = 0.92 \delta_s^2.
\]

Since the estimate of \( \delta_s^2 \) in the case of 154 day weight is less than zero, it must be assumed that the \( g_i \) for all the lines are equal with respect to this measurement. The adjusted values of the other \( g_i \) when added to \( \hat{\mu} \) give the values shown in Table 36.

Table 36

<table>
<thead>
<tr>
<th>Values of ( \hat{\mu} + \hat{g}_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Litter Number</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Line</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>H</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>J</td>
</tr>
<tr>
<td>K</td>
</tr>
<tr>
<td>S</td>
</tr>
</tbody>
</table>

Two facts are clear from a comparison of the \( \hat{\mu} + \hat{g}_i \) of Table 18 with the \( \hat{\mu} + \hat{g}_i \) of Table 36. First, the variation in general combining abilities among lines appears much reduced when \( \hat{\mu} + \hat{g}_i \) is used as the basis for appraisal of lines. Second, decisions with respect to keeping or discarding
lines would be essentially the same whether one uses the least squares estimates or the adjusted least squares estimates. In a completely balanced design the ranks of the lines would, in fact, be identical for the two methods. In a design as unbalanced as the one in this study, only minor changes in rank occur. If it is a question, however, of selecting specific crosses for future use, the ranks might be quite different, depending upon whether one uses as a criterion \( \hat{\mu} + \hat{\epsilon}_i + \hat{\epsilon}_j + \hat{\sigma}_{ij} \) or an index, \( I = b_1 \hat{\epsilon}_i + b_2 \hat{\epsilon}_j + b_3 \hat{\sigma}_{ij} \) since the index emphasizes the \( \hat{\epsilon}_i \) and the \( \hat{\sigma}_{ij} \) according to the amount of information available concerning each of them and according to the relative magnitudes of \( \sigma_g^2 \), \( \sigma_s^2 \), and \( \sigma_e^2 \). The computation of individual indexes for each of the possible crosses scarcely seems justified in this study since the estimates required in constructing such indexes have relatively large sampling errors. A rather good approximation to these indexes can, however, be obtained by assuming that the design was a balanced one. In that case, the appropriate index is \( I = b_1 (\hat{\epsilon}_i + \hat{\epsilon}_j) + b_2 \hat{\sigma}_{ij} \), where

\[
b_1 = \frac{(p-2) \sigma_g^2 + \sigma_s^2}{(p-2) \sigma_g^2 + \sigma_s^2 + \sigma_e^2}
\]

\[
b_2 = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2}
\]

Utilizing the estimates of \( \sigma_g^2 \), \( \sigma_s^2 \), and \( \sigma_e^2 \) presented in Table 34 (except \( \sigma_g^2 \) assumed = 0 in 154 day litter weight), the following coefficients were obtained:
<table>
<thead>
<tr>
<th>Category</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number at 0</td>
<td>.32</td>
<td>.29</td>
</tr>
<tr>
<td>Number at 21</td>
<td>.49</td>
<td>.16</td>
</tr>
<tr>
<td>Number at 56</td>
<td>.45</td>
<td>.22</td>
</tr>
<tr>
<td>Number at 154</td>
<td>.38</td>
<td>.23</td>
</tr>
<tr>
<td>Weight at 0</td>
<td>.55</td>
<td>.38</td>
</tr>
<tr>
<td>Weight at 21</td>
<td>.53</td>
<td>.27</td>
</tr>
<tr>
<td>Weight at 56</td>
<td>.42</td>
<td>.22</td>
</tr>
<tr>
<td>Weight at 154</td>
<td>.36</td>
<td>.36</td>
</tr>
</tbody>
</table>
IV. DISCUSSION AND APPLICATIONS

This study has shown that single crosses among inbred lines can furnish several types of information including: (1) estimates of the general combining abilities and maternal abilities of the lines, (2) estimates of the values of specific crosses, (3) estimates of the relative importance of general, maternal, specific, and sex linkage effects as causes of variation among single crosses, and (4) an estimate of the genetic correlation between additive genetic and maternal abilities of the lines. The methods presented in Section II describe how such information can be obtained from data which are unbalanced with respect to number of experimental units per cross and with respect to such factors as age and inbreeding of the dams. The estimates obtainable from single crosses can be used as follows: (1) to select lines for top crossing, (2) to select lines for use as the female parent, (3) to select specific crosses for further testing or for commercial use, (4) to study the value of the estimates of general, maternal, and specific effects in predicting the outcome of three-way and four-way crosses, (5) to study the value of information on the performance of an inbred line in predicting the performance of the line as line of sire and as line of dam in single crosses, and (6) to plan most efficient types of testing programs. If the information from single cross tests is to be properly utilized for the 6 purposes stated above, selection indexes must be constructed in which the independent variables are rather complex least squares estimates.

A. Relation of the Methods of Analysis of Single Crosses to More Familiar Methods in Animal Breeding

The methods utilized in this study differ not at all in principle from the commonly used methods in animal breeding. As an example of a common
problem, let us assume that we wish to choose from a group of animals the
one with the highest probable breeding value and that the information avail-
able for estimating breeding value includes one record on the individual
and one record on its dam. Assuming that all records were made in the same
herd, that there have been no important time trends in management, and that
genotype and environment are uncorrelated, the appropriate index on which
to base the selection is

\[ I = b_1 x_1 + b_2 x_2, \]

where \( x_1 \) is the record on the indi-
dividual and \( x_2 \) is the record on his dam. The \( b \)'s come from the solution
to these equations:

\[
\begin{align*}
\sigma^2_{x_1} & \quad b_1 + b_2 & \quad \sigma_{x_1 x_2} = \sigma_{x_1 G} \\
\sigma^2_{x_2} & \quad b_1 \sigma_{x_1 x_2} & \quad b_2 \sigma^2_{x_2} = \sigma_{x_2 G},
\end{align*}
\]

where \( G \) is the breeding value of the individual. The variances and co-
variances needed to construct the index are ordinarily estimated from an
analysis of variance or a regression analysis. The manner in which the es-
timates of the variances is made depends on the assumptions of the model
(path coefficient diagram for example) which is thought to fit the biology
of the material. Each step in the analysis of the single cross test is
analogous to one of the steps in the simple example of constructing a se-
lection index to select for breeding value on the basis of own performance
and parent's performance. Thus, the least squares estimates, \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \),
correspond to \( x_1 \) and \( x_2 \). \( \hat{\sigma}^2_{\alpha_1} \) corresponds to \( \sigma^2_{x_1} \), \( \hat{\sigma}_{\alpha_1 \alpha_2} \) corresponds to
\( \sigma_{x_1 x_2} \), and \( \hat{\sigma}_{\alpha_1 \alpha_2, T} \) corresponds to \( \sigma_{x_1 G} \). The estimate of \( \sigma^2_\alpha \) from the least
squares sums of squares corresponds to the estimate of heritability needed
in the simple example above. The mathematical model, \( y_{ijk} = \mu + \alpha_i + \alpha_j + 
\alpha_{ij} + r_{ij} + e_{ijk} \), corresponds to the path coefficient diagram of the selection
problem.
The relationship between the methods of this study and conventional methods in animal breeding can also be illustrated by a statement of certain facts concerning the latter. Single records or means of several records are, in fact, least squares estimates, ordinary analysis of variance and regression analysis are actually least squares analyses, and the estimation of variances from an analysis of variance involves taking expected values of least squares sums of squares. The relative complexity of the statistical methods used in the analysis of single cross data is due largely to the non-orthogonality of the single cross data and to a lesser extent to the fact that maternal, specific, and sex linkage effects as well as additive genetic effects are taken into account.

B. Interpretation of the Experimental Results

The most striking results of the analysis of the single cross data available for this study are the apparent small differences among lines in general combining ability, the lack of any evidence of maternal differences among lines, and the relatively large differences among specific effects. Since this is the first study in which these three sources of variance have been estimated from single crosses among inbred lines of swine, the results cannot be compared directly with any previous results. It is possible, on the basis of genetic theory to predict the additive genetic variance among lines if it is assumed that the genetic variance in the original population from which the lines were formed is known and that there has been no selection either within lines or among lines. Previous heritability estimates may therefore be compared with the results of this study. On the other hand, there seem to be no previous estimates in outbred populations of swine of
maternal effects which are comparable to the maternal effects as defined in this study. Maternal differences as defined here include both pre-natal and post-natal effects and are due entirely to differences among lines in genes influencing mothering ability. In previous studies of maternal effects it has not been possible to separate genetic differences with respect to mothering ability of the sow from environmental factors common to litter mates. (Lush, Hetzer, and Culbertson, 1934.) The reason that genetic differences among lines in their maternal abilities can be isolated in crosses among inbred lines is the fact that each line is used both as the male parent and as the female parent, and consequently, the maternal effect can be estimated by comparing the performance of each line with respect to these two characteristics. The post-natal portion of maternal differences could be determined in outbred populations by an experimental design in which members of litters were switched from one sow to another if it were not for the difficulty in accomplishing such interchanges. It would also seem possible to obtain estimates of combined pre-natal and post-natal maternal effects by crossing reciprocally different families, say full sib families. Then the analysis would be identical with that described for single crosses among inbred lines, the families corresponding to the lines.

Two factors preclude the possibility of predicting the size of the variance for specific combining ability. (1) No estimates are available concerning the magnitude of epistatic and dominance deviations in outbred populations of swine. (2) Even if there were such estimates, there is no way of stating precisely what effect inbreeding has on the size of these deviations and on their distribution among lines and within lines. Studies which have been made of "nicking" effects in outbred populations include those of Seath and
Lush (1940) on milk and butterfat production in dairy cattle, of Lerner (1945) on sexual maturity in poultry, and of Hazel and Lamoreux (1947) on sexual maturity and body weight in poultry. All of these studies indicated that there were small "nicking" effects. The only previous study on "nicking" effects among crosses of inbred lines is that of Sprague and Tatum (1942) on yields of single crosses of corn. They showed that specific effects were important in crosses among unselected lines.

1. General combining ability

The types of evidence which point to the apparent small differences among the 12 inbred lines of swine are (1) the results of the tests of significance of $G$, (2) the comparisons of the least squares estimates of the general combining abilities for the various lines with the sampling errors of the estimates, and (3) the estimates obtained for the general combining ability variance components. The mean square for $G$ was significant at the 5 per cent level in only 3 of the 8 measurements when $\sigma_e^2$ was used as the error mean square. When the mean square for specific combining ability was used as the error term, none of the mean squares for $G$ was significant. The least squares estimates for the different line effects do appear quite different from one line to another (Table 18), but these estimates are subject to large sampling errors, and when the estimates are adjusted in accordance with the assumption that the lines are randomly drawn from some population of lines, the differences among them are small indeed (Table 36). The estimates of $\sigma_g^2$ range from 0 to 3 per cent of the estimate of $\sigma_e^2$ for the 8 measurements. As was stated previously, $\sigma_g^2$ equals one-fourth the additive genetic differences among lines. Consequently, if the estimates of
\( \sigma_g^2 \) are multiplied by 4, estimates of the additive genetic differences among lines do not exceed 12 per cent of the error variance.

If \( \sigma_g^2 \) is actually less than should be expected on the basis of previous heritability estimates, two possible reasons might be suggested. First, the foundation stock used in establishing the lines may have been closely related or very similar genetically. Second, intensive selection in the same direction in the development of all lines may have made them very similar genetically. Actually, six of the lines trace to common foundation stock bred as a closed herd for 7 or 8 years before the lines were separated. One of the other lines traces in part to this same stock. It seems reasonable that the 12 lines differ less than would 12 lines developed from stock selected at random from the Poland China breed. The second reason is a less plausible one for it seems improbable that selection for large litter size and rapid growth rates in all lines could have been very effective in eliminating differences among lines. Studies are now in progress at the Iowa Agricultural Experiment Station to determine the amount of selection actually practiced in developing the lines.

If there had been no selection either among or within lines, and if the foundation stock had been chosen at random, the expected variance among lines would equal 2f \( \sigma_{g_0}^2 \), where \( \sigma_{g_0}^2 \) equals the additive genetic variance of the population from which the foundation stock came and f equals the inbreeding of the progeny of matings within lines. Assuming for the present that these assumptions are correct, the estimates of \( \sigma_g^2 \) obtained in this study can be utilized to estimate heritability in the original population. Assuming that \( f = .34 \), the expectation of \( \bar{\sigma}^2 = 2 (.34) \sigma_{g_0}^2 \), where \( \bar{\sigma}^2 \) is the additive genetic variance among inbred lines. Then since \( \bar{\sigma}^2 = 4 \sigma_g^2 \),
\[ 4 \sigma_g^2 = 0.68 \sigma_{a0}^2, \text{ and} \]
\[ \sigma_g^2 = 0.17 \sigma_{a0}^2. \]

But since \( \sigma_{a0}^2 = \text{heritability} \times \sigma_{T0}^2 \), where \( \sigma_{T0}^2 \) is the phenotypic variance in the original population, an estimate of heritability in the original population is \( \frac{\sigma_g^2}{0.17 \sigma_{T0}^2} \). A slight under-estimate of \( \sigma_{T0}^2 \) can be obtained from the mean square for \( R(T) - R(\mu, a, F) \). Utilizing these mean squares and the estimates of \( \sigma_g^2 \) reported in Table 34, the following estimates of heritability were obtained:

- Number at birth = 0.01
- Number at 21 days = 0.13
- Number at 56 days = 0.09
- Number at 154 days = 0.05
- Litter weight at birth = 0.09
- Litter weight at 21 days = 0.12
- Litter weight at 56 days = 0.07
- Litter weight at 154 days = 0.

When these estimates of the heritability of litter characteristics are compared with estimates previously reported (Phillips, 1947), it is not at all certain that the general combining abilities of the lines actually differ less than should be expected. Common relationship among some lines and similar standards of selection both would tend to underestimate the genetic variability in the original population. Furthermore, the estimates of \( \sigma_g^2 \) come from a mean square (among G) with only 11 degrees of freedom, and variances based on so few degrees of freedom are rather poorly estimated, regardless of the number of observations within each of the classes. Previous
heritability estimates are also subject to large sampling errors. Thus, there are good reasons for expecting considerable discrepancy between estimates of heritability obtained from the line crosses and estimates obtained from other material or made by other methods.

The small differences found among lines in this study for characteristics such as litter number and litter weight do indicate that there will be many errors made in selecting among lines unless the lines have been very thoroughly tested. This does not mean, however, that most rapid genetic progress can be made by very carefully testing lines, since doing the testing this carefully probably would preclude the alternative policy of making more lines and testing each of them less carefully. A few examples of the optimum number of lines to make and test are presented later in this section.

2. Maternal effects

There seems little doubt that in this sample of single crosses maternal effects are very small, if present at all. The evidence for this is that the mean squares for M were less than those for error in all measurements except 21 day and 154 day litter weights and in the latter traits the F value was much less than that needed for significance at the 5 percent level. Also, the estimates of $\sigma_m^2$ coming from the least squares analysis were negative for all measurements except 21 day weight. The fact that maternal differences were non-existent or at least very small made it possible to obtain better estimates of the general and specific combining abilities than would otherwise have been the case, since the general combining ability could, therefore, be estimated from both the line
of sire and the line of dam. The absence of any evidence of maternal differences among lines is quite surprising, but there seems to be no way of comparing this result with previous studies on maternal differences. This is apparently the first study in which genetic differences could be separated from environmental factors influencing the mothering ability of sows. The same factors suggested as possible causes of the small differences in general combining ability among the lines may have made the variances due to maternal differences small also. Furthermore, the estimate of $\sigma_m^2$ is subject, as is the estimate of $\sigma_g^2$, to large sampling errors due to the fact that both are estimated from mean squares with only 11 degrees of freedom.

If there had been any evidence that maternal ability differed from line to line, an estimate of the genetic correlation between general combining ability and maternal ability would have been made, for this is a matter of rather considerable theoretical and practical interest. The method by which such an estimate could have been made was described in Section II G. As an example of the usefulness of such an estimate, let us assume that in a particular population of lines the genetic correlation between general combining ability and maternal ability is a large negative one. Then since performance as line of dam is equal to $g_1 + m_1$, the variance of the differences among the lines in abilities as lines of dam = $\sigma_g^2 + \sigma_m^2 + 2\sigma_{gm}$, and if $\sigma_{gm}$ is large and negative, the value of the variance among lines of dam might be small. Consequently, selection among lines for performance as line of dam could be quite ineffective if selection were based entirely on performance as line of dam. If, however, selection is based on an index, $I = b_1\hat{g}_1 + b_2\hat{m}_1$, and this index is constructed as
described in Section II C 6 b, selection could be considerably more effective.

3. Specific Effects

The results of this study indicate that there are real differences in specific combining abilities in this population of lines. That is, specific crosses are more unlike than would be expected under the hypothesis that differences among crosses are due entirely to differences in the average effects of lines, to maternal differences among lines, and to sampling error (including errors of Mendelian sampling). Although only three of the measurements showed a significant ($P < .05$) $F$ value in the test of the hypothesis that $s_{ij} = 0$, the other five had $F$ values only slightly less than the value required for significance at the 5 per cent level. Furthermore, the estimates of $\hat{\sigma}_s^2$ for all measurements were large in relation to the estimates of general and maternal effects. In fact, the estimates seem extraordinarily high, one of them being almost one-fifth as large as $\hat{\sigma}_e^2$.

A reasonable amount of confidence can, however, be placed in these estimates since they are unbiased and since they are obtained from mean squares with 30 degrees of freedom.

This large specific combining ability effect seems all the more surprising when it is considered that the 12 lines were formed from the same breed and that several of the lines are related. Studies of other populations of lines are needed, however, to determine what specific differences can reasonably be expected.

If further studies should confirm that large differences in specific combining ability exist in other populations of lines, there are some important applications from the standpoint of making and testing inbred lines.
The consequences of varying $\sigma^2_6$ on appropriate designs for testing inbred lines are described in Section IV D. If specific effects are really as large as the estimates found in this study, it will be advantageous to advance the inbreeding further before starting a testing program than was suggested by Dickerson (1942), who considered general combining ability only. It would seem desirable also to estimate the importance of specific combining ability in random bred populations. This can be done by a system of diallel or polyallel matings. The methods described in Section II for obtaining estimates of variance components could be utilized to estimate the variance for specific effects in such matings, even though the design were not a perfectly balanced one.

4. Sex linkage effects

It is not at all surprising that there was no evidence of differences among crosses due to sex linkage. Since the pig has 20 pairs of chromosomes, the amount of data in this study could hardly be expected to provide a measure of such differences, unless the sex chromosome carried a disproportionately high fraction of the genes which affect the characters studied here.

C. Utilization of the Estimates of Line and Cross Line Characteristics

1. Selection of lines for top crossing

The proper method for utilizing the individual estimates of line and cross line characteristics in this or other studies of single crosses depends upon what use is to be made of the lines. Let us first assume that, on the basis of the single cross test alone, lines are to be selected for
use in top crossing. In that case selection is entirely for general combining ability. If the true values of the $g_4$ for all lines were known, the lines would be selected on the basis of that knowledge. What we have available for estimating the $g_4$ are the least squares estimates of them. The expectation of each of these estimates contains not only $g_4$ but also fractions of each of the other $g_k$ of the parameters for specific effects, and of errors. Consequently, if we are to obtain the best estimate of the general combining ability of the line, we must regress the least squares estimate toward 0 by an amount dependent upon the sizes of the various coefficients in the expectation and on the basis of the estimates of general, specific, and error variances. The method for doing this was described in Section II, and the regressed estimates were presented in Table 36. A question arises concerning what to do regarding 154 day litter weight, for which the estimate of $\sigma_g^2$ is less than 0. If this variance is assumed equal to zero and is then substituted in the regression formula, all of the estimates of $g_4$ become 0. Consequently, it would be assumed that the lines are of equal value for top crossing. Even though there may be good reasons for expecting the variance to be small, there is no good reason for thinking that it is zero. Consequently, it would seem to make better sense to assume that the negative estimate was due to accidents of sampling and to substitute for it some reasonable guess as to the value of the variance.

2. Selection of lines for use as female parents

If the results of the single cross tests are to be used to select lines for use as the female parent in matings with boars from stock which has not previously been tested in matings with the tested lines, selection would be
on the basis of the index \( I = b_1 \hat{g}_1 + b_2 \hat{m}_1 \). If one accepts the hypothesis that \( \sigma_m^2 = 0 \), the index reduces to exactly the same one as when selection is for lines to use in top crossing. On the other hand, one might prefer to assume that there actually are small differences in maternal abilities among lines and to utilize some estimated value for this variance in constructing the index.

3. Selection of specific crosses

If the single cross test results are to be used to select specific single crosses for commercial utilization or for further testing, the index \( I_1 = b_1 \hat{g}_1 + b_2 \hat{e}_j + b_3 \hat{g}_j + b_4 \hat{s}_{ij} + b_5 \hat{r}_{ij} \), should be utilized. If one assumes, as the data indicate, that \( \sigma_m^2 \) and \( \sigma_r^2 = 0 \), the appropriate indexes are approximately the ones given in Section III. The relative efficiency of selection based on this index can be compared with selection based entirely on the performance of the various specific crosses or on the performance of specific crosses regressed in accordance with the relative sizes of the variances for error and for differences among specific crosses. The correlation for each selection method between the index and the true value of the cross determines their relative efficiencies. Assuming that \( \sigma_m^2 = \sigma_r^2 = 0 \), the true value of a specific cross is \( \hat{g}_1 + \hat{g}_j + \hat{s}_{ij} \). Then in a balanced design

\[
\begin{align*}
\delta_{T1}^2 &= \frac{\sigma_{I_1}^2}{\delta_T^2} \\
\sigma_{I_1}^2 &= \frac{2}{p} \left[ (p-2) \delta_g^2 + \delta_s^2 \right] \\
\delta_T^2 &= 2 \delta_g^2 + \delta_s^2
\end{align*}
\]

If we assume that the design was a balanced one, the second method of appraisal of specific crosses is equivalent to selection based solely on \( \hat{y}_{1j} \).
the mean of the observations of the mating, male of $i^{th}$ line by female of the $j^{th}$ line. Assuming the mean is known,

$$\delta (T I_2) = 2 \delta g^2 + \delta s^2$$

$$\delta I_2^2 = 2 \delta g^2 + \delta s^2 + \frac{1}{n} \delta \epsilon^2,$$

where $n$ is the number of experimental units per reciprocal cross

$$\delta T^2 = 2 \delta g^2 + \delta s^2$$

Now if it is assumed that $\delta g^2 = 2$, $\delta s^2 = 4$, $\delta \epsilon^2 = 100$, $p = 12$, and $n = 2$, the expected correlation between the index and the true value is $.55$ for the most efficient index and $.37$ for the less efficient one. If, instead of the above numerical values for the variances, $\delta s^2$ is assumed equal to 12, a value more nearly in line with results of this study, the expected correlation between the true value and the index is $.62$ for the efficient index and $.49$ for the inefficient one. These estimates illustrate that in certain selection problems utilization of the more powerful statistical methods may materially increase the effectiveness of selection at the cost only of some computational labor.

D. Choice of an Efficient Testing Program

The progress to be expected through use of an index constructed as described in Section II is $\bar{I}_n$, where $\delta I$ is the standard deviation of the properly constructed index and $\bar{I}_n$ is the number of standard deviations by which the mean index of the selected group exceeds the mean of the population. Consequently, the estimates of variances obtained in this study can be used to study the expected progress from different types of testing programs.
1. **Value of further testing of existing lines**

Let us first consider how much gain can be expected from further testing of the twelve lines, subsequent testing to be done in such a way as to effect a completely balanced design. Let us first assume that we are interested in selecting a single line with highest $g_1$, the goal of selection if the line were to be used for top crossing. In this case, the index is

$$I = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_s^2 + \sigma_e^2} \cdot \hat{g}_1 .$$

Then

$$V(I) = \frac{\frac{p-1}{p} \sigma_g^4}{\sigma_g^2 + \sigma_s^2 + \sigma_e^2}$$

Assuming that $\sigma_g^2 = 2$, $\sigma_s^2 = 8$, and $\sigma_e^2 = 100$, the value of $V(I)$ is

$$\frac{11}{8.4 + 15} .$$

As before, $n$ is the number of litters per reciprocal cross.

Since $\bar{x}_n$ is constant when the number of lines is constant, the expected progress is directly related to $\sigma_I$. The values of $\sigma_I$ for various $n$ are,

1 = .47  
2 = .69  
3 = .82  
4 = .91

Expressing the expected progress as a fraction of the possible maximum $(n = \infty)$, the following values for different $n$ are obtained.

1 = .36  
2 = .53  
3 = .63  
4 = .69  
5 = .73  
10 = .85
It is obvious from the data above that the approach to maximum progress through increasing \( n \) is very slow indeed. Furthermore, each increase of 1 in \( n \) means an increase of 132 litters tested.

If the object of the single cross test is selection of the best possible specific cross, the appropriate index is

\[
I = b_1 \hat{e}_1 + b_2 \hat{e}_j + b_3 \hat{e}_{ij},
\]

where

\[
b_1 = b_2 = \frac{(p-2) \sigma^2_g + \sigma^2_s}{\sigma^2_g + \sigma^2_s + \sigma^2_e} \quad \text{and}
\]

\[
b_3 = \frac{\sigma^2_s}{\sigma^2_s + \sigma^2_e}. \tag{1}
\]

The variance of this index is

\[
\frac{(p-2) \sigma^2_g + \sigma^2_s}{p \sigma^2_g + \sigma^2_s + \sigma^2_e} + \frac{p-1}{p-3} \frac{\sigma^4_s}{\sigma^2_s + \sigma^2_e}. \tag{2}
\]

Substituting as before the numerical values for the variances and for \( p \), the following values are obtained for \( \sigma_I \) when \( n \) varies as shown below.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \sigma_I )</th>
<th>( n )</th>
<th>( \sigma_I )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.61</td>
<td>5</td>
<td>2.52</td>
</tr>
<tr>
<td>2</td>
<td>2.01</td>
<td>10</td>
<td>2.83</td>
</tr>
<tr>
<td>3</td>
<td>2.25</td>
<td>\infty</td>
<td>3.35</td>
</tr>
<tr>
<td>4</td>
<td>2.40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The expected progress expressed as a fraction of the maximum possible (\( n = \infty \)) is as follows:
2. Determining the optimum number of lines to test

An important question is whether it would be more profitable to make and test more lines rather than to continue testing the existing lines. Some idea as to the optimum type of testing program can be obtained by assuming that there is a fixed number of litters which can be tested and by then choosing \( p \) so as to maximize the expected progress. Let us assume that 500 tests can be made but that the number of lines can be altered at will. Let us also assume that \( \sigma^2 = 100, \sigma^2_g = 2, \sigma^2_r = \sigma^2_m = 0, \) and that \( \sigma_s^2 \) ranges from 0 to 20. Different values of \( \sigma_s^2 \) are used to illustrate the effect of varying levels of that variance on the optimum testing program and upon the progress which can be expected. As before, expected progress = \( \sigma \bar{x}_n \), and values of the expectations for differing \( p \) and \( \sigma_s^2 \) are presented in Table 37. The values of \( \bar{x}_n \) were obtained from Tippett's (1925) tabulation of the expected value of the range of \( n \) observations from a normal distribution with unit variance. For example, the approximate value of \( \bar{x}_n \) for \( p = 5 \) is 1.54, the expected value of the largest observation in a sample of \( \frac{p(p-1)}{2} = 10 \) observations.

This table shows that the expected value of the cross with the highest index increases with increasing values of \( \sigma_s^2 \) no matter what is the value of \( p \). This is expected, since the chance of selecting a cross which has a high specific value increases as \( \sigma_s^2 \) increases. The optimum number of lines to test is influenced by the size of \( \sigma_s^2 \) also. When \( \sigma_s^2 \) is assumed

\[
\begin{align*}
1 &= .48 \\
2 &= .60 \\
3 &= .67 \\
4 &= .72 \\
5 &= .75 \\
10 &= .85
\end{align*}
\]
to be zero, the optimum number is \( p = 23 \); that is, \( n = 1 \). When \( \sigma_s^2 \) is assumed to equal 4, the optimum number of lines is 16, or \( n = 2 \). It might seem logical that more lines should be tested when \( \sigma_s^2 \) is positive than when it is zero, since more specific crosses could thereby be tried and the chances for obtaining a superior one increased. The fact is, however, that when differences in specific combining ability do exist, a reasonable number of litters of each of the crosses must be tested in order to obtain much information concerning the value of a specific cross above what can be obtained from knowledge of only the \( \hat{g}_i \). When \( \sigma_s^2 = 8 \), the optimum number of lines to test is either 9 or 11, a number still smaller than the optimum when \( \sigma_s^2 = 4 \). The logical reason for this apparently is that increasing emphasis should now be placed on specific combining ability, and consequently, it is desirable to obtain still more information about the value of each specific cross. A point is reached, however, when \( \sigma_s^2 \) becomes large enough to cause an increase in the optimum number of lines tested. For example, when \( \sigma_s^2 = 12 \), the optimum number of lines is 11, and when \( \sigma_s^2 = 20 \), the optimum number is 13.

Table 37

Expected Progress from Selecting the Single Cross With Highest Index in a Balanced Single Cross Test of 500 Litters

<table>
<thead>
<tr>
<th>( p )</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>12</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.85</td>
<td>1.70</td>
<td>2.11</td>
<td>2.51</td>
<td>3.18</td>
</tr>
<tr>
<td>5</td>
<td>2.06</td>
<td>2.91</td>
<td>4.39</td>
<td>5.25</td>
<td>6.67</td>
</tr>
<tr>
<td>7</td>
<td>2.67</td>
<td>3.38</td>
<td>5.21</td>
<td>6.30</td>
<td>8.10</td>
</tr>
<tr>
<td>9</td>
<td>3.06</td>
<td>3.70</td>
<td>5.61</td>
<td>6.83</td>
<td>8.89</td>
</tr>
<tr>
<td>11</td>
<td>3.27</td>
<td>3.80</td>
<td>5.61</td>
<td>6.87</td>
<td>9.07</td>
</tr>
<tr>
<td>13</td>
<td>3.41</td>
<td>3.90</td>
<td>5.57</td>
<td>6.82</td>
<td>9.08</td>
</tr>
<tr>
<td>16</td>
<td>3.52</td>
<td>3.93</td>
<td>5.40</td>
<td>6.60</td>
<td>8.87</td>
</tr>
<tr>
<td>23</td>
<td>3.64</td>
<td>3.86</td>
<td>5.02</td>
<td>5.98</td>
<td>8.04</td>
</tr>
</tbody>
</table>
3. **Top cross vs. single cross test**

One way of comparing the relative efficiencies of single cross tests and top cross tests is to compare the expected value of the best cross selected on the basis of top cross tests with the expected value of the best cross selected on the basis of single cross tests. If the selection were to be made on the basis of a top cross test, the procedure would be to select the two lines with the highest \( \hat{g}_1 \) on the top cross test. The index for predicting the performance of the \( ij^{th} \) cross is

\[
I = b(\hat{g}_i + \hat{g}_j) = \frac{\sum g_i^2}{\sum g_i^2 + \sum g_n^2} (\hat{g}_i + \hat{g}_j)
\]

where \( C \sum g_t^2 \) is the portion of \( \sum g_i^2 \) due to line x tester interaction and \( n \) is the number of litters of each top cross. Since we have no estimate of \( \sum g_t^2 \) it will be assumed equal to zero. This assumption will make the top cross test appear more efficient than it actually is. The variance of this index is

\[
\frac{2 \sum g_i^2}{\sum g_i^2 + \sum g_n^2} \cdot \frac{k^2}{n^2}
\]

\( k = 100 \), and \( k = 500 \), the expected value of the cross with the highest index for differing values of \( p \) is as follows:

<table>
<thead>
<tr>
<th>( p )</th>
<th>( I )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.85</td>
</tr>
<tr>
<td>11</td>
<td>2.86</td>
</tr>
<tr>
<td>5</td>
<td>1.94</td>
</tr>
<tr>
<td>13</td>
<td>2.93</td>
</tr>
<tr>
<td>7</td>
<td>2.44</td>
</tr>
<tr>
<td>16</td>
<td>2.98</td>
</tr>
<tr>
<td>9</td>
<td>2.74</td>
</tr>
<tr>
<td>23</td>
<td>2.96</td>
</tr>
</tbody>
</table>

Comparing the progress to be made by the top cross method with that by single cross tests when \( \sum g_i^2 = 0 \) (second column of Table 37), progress by the single cross method is the greater, the reason being that \( g_i \) is estimated from both line of sire and line of dam in the single cross test.
When $\sigma^2_e$ is greater than zero, the relative advantage of the single cross test increases with increasing values of $\sigma^2_e$.

4. Early culling of lines followed by further testing

The most efficient testing program is undoubtedly one in which a certain number of lines are tested the first year, a certain number are discarded on the basis of this first test and the remainder are tested a second year, and so on until a line or cross is chosen which has the highest index for the tests during all years. The computations required for predicting the optimum program are very burdensome since a very large number of testing programs are possible, and a maximum cannot be calculated by the usual methods of differential calculus. The following example is given to illustrate the general procedure for estimating what proportion of lines to discard at the end of a testing year. Assume that there are 12 lines to be tested and that 100 litters can be tested each year. Assume that there is to be a two year top cross testing program and that at the end of the second year the line appearing the best in top cross performance is to be selected for future use. Assume further that $\sigma^2_g = 1$, $\sigma^2_e = 50$, and $\sigma^2_{gt} = 0$. The index for estimating $g_1$ is

$$I = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_e}{n}} \hat{g}_1.$$ 

The variance of the index is

$$V(I) = \frac{p-1}{p} \frac{\sigma^4_g}{\sigma^2_g + \frac{\sigma^2_e}{n}}.$$ 

The expected value of the mean of the $g_1$ for the $q$ lines with the highest $\hat{g}_1$ is $\sigma_{1xq}$, where $\bar{x}_q$ is the mean of the $q$ highest deviates from a normal distribution with unit variance. These values can be obtained from the
tables of Fisher and Yates (1948). Then $\sigma_{g_1}^2$ is computed by the method described by Dickerson and Hazel (1944), $\sigma_{g_1}^2$ being the variance among the $q$ selected lines. Finally, the expected value of $g_1$ for the line having the highest $\hat{g}_1$ in the two tests is computed, using the index

$$I = \frac{\sigma_{g_1}^2}{\sigma_{g_1}^2 + \sigma_e^2} \left( \frac{\hat{g}_1 - \bar{X}_q}{\sigma_f} \right).$$

In the present example it turns out that the expected value of the line chosen on the basis of a two year test with differing values of $q$ is as follows:

<table>
<thead>
<tr>
<th>$q$</th>
<th>Expected Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.47</td>
</tr>
<tr>
<td>3</td>
<td>.61</td>
</tr>
<tr>
<td>4</td>
<td>.67</td>
</tr>
<tr>
<td>5</td>
<td>.71</td>
</tr>
<tr>
<td>6</td>
<td>.73</td>
</tr>
<tr>
<td>7</td>
<td>.74</td>
</tr>
<tr>
<td>8</td>
<td>.75</td>
</tr>
<tr>
<td>9</td>
<td>.76</td>
</tr>
<tr>
<td>10</td>
<td>.77</td>
</tr>
<tr>
<td>11</td>
<td>.77</td>
</tr>
<tr>
<td>12</td>
<td>.78</td>
</tr>
</tbody>
</table>

Therefore, progress is a maximum when all of the lines are tested for two years. Not much is lost, however, by culling approximately half of the lines at the end of the first year, and such culling might release facilities for expanding the remaining lines or for making additional lines.

**E. Further Applications of the Method of Selecting Simultaneously for Both General and Specific Effects**

The method described for constructing an index to select the best specific cross on the basis of a single cross experiment has more general applications such as selection for specific combining ability in outbred populations and in breed crosses. For example, let us assume that semen
from a number of bulls has been used on a relatively random sample of dairy herds in a state. Then the production records of the daughters of these bulls can be used to estimate the general combining ability of each bull, an effect for each of the farms (this effect including both genetic and environmental differences), effects peculiar to the progeny of a particular bull used on a particular farm, the variance due to bull differences, the variance due to farm differences, the variance due to the interaction between farms and bulls, and error variance. On the basis of the estimates of these parameters, separate indexes can be constructed for each farm and these indexes then used to select semen for that particular farm. The indexes would, of course, differ from farm to farm if bull by farm interaction is real. Assuming that the mathematical model describing the production of a particular cow is

\[ y_{ijk} = \mu + b_i + f_j + bf_{ij} + e_{ijk}, \]  

where \( b_i \) is the sire effect, \( f_j \) is the farm effect, \( bf_{ij} \) is the specific sire by farm effect, and \( e_{ijk} \) is an effect peculiar to a particular cow, the appropriate index for selecting bulls for use on a particular farm is

\[ I = a_1 \hat{b}_i + a_2 \hat{bf}_{ij} \]

The equations needed to construct the index for the \( i^{th} \) bull on the \( j^{th} \) farm are

\[ a_1 \hat{b}_i^2 + a_2 \hat{bf}_{ij} = \hat{b}_i (b_i + bf_{ij}) \]
\[ a_1 \hat{b}_i \hat{bf}_{ij} + a_2 \hat{bf}_{ij}^2 = \hat{bf}_{ij} (b_i + bf_{ij}) \]

It seems rather certain that the \( a_2 \) coefficient would be small in comparison to the \( a_1 \) because of the difficulty of estimating the \( bf_{ij} \) accurately. If, however, \( \sigma_{bf}^2 \) should happen to be large in comparison to \( \sigma_b^2 \), \( a_2 \) might
be large enough to make it worthwhile to consider $\hat{b}_{1j}$ as well as $\hat{b}_1$ in selecting a bull for use on a particular farm. At any rate, the use of such an index would make maximum use of information on both the additive genetic value of a bull and the specific value of that bull for a particular farm. Selection of a male for mating with a particular family of females could also be done in the same manner as selection of a bull for use on a particular farm; the only change needed in the model is the substitution of families for farms.

The index method for selecting specific and general effects can also be extended to selection of a breed cross. The breed cross with the highest probability of future success would not necessarily be the particular cross which performed best in a test of breed crosses. Rather the proper choice is the cross with highest index, $I = b_1 \hat{m}_i + b_2 \hat{f}_j + b_3 \hat{m}_i \hat{f}_j$, where $\hat{m}_i$ is the performance of the $i^{th}$ breed as the male parent, $\hat{f}_j$ = the performance of the $j^{th}$ breed as the female parent, and $\hat{m}_i \hat{f}_j$ is the amount by which the cross of the $i^{th}$ by the $j^{th}$ breed exceeds the expectation based on the values of $\hat{m}_i$ and $\hat{f}_j$. The index would be constructed from the solution of the following set of equations:

$$b_1 \hat{m}_i^2 + b_2 \hat{m}_i \hat{f}_j + b_3 \hat{m}_i \hat{f}_j = \hat{m}_i^2 (m_i + f_j + mf_{ij})$$
$$b_1 \hat{m}_i \hat{f}_j + b_2 \hat{f}_j^2 + b_3 \hat{f}_j \hat{m}_i = \hat{f}_j^2 (m_i + m_j + mf_{ij})$$
$$b_1 \hat{m}_i \hat{f}_j + b_2 \hat{f}_j \hat{m}_i + b_3 \hat{m}_i^2 = \hat{m}_i \hat{f}_j (m_i + m_j + mf_{ij})$$

If $\hat{m}_i$, estimated from analysis of top cross tests, is substituted for $\hat{m}_i$ in the above example, an index can be constructed for selecting lines to use in top crosses on particular breeds.
The results of single crosses among 12 inbred Poland China lines at the Iowa Agricultural Experiment Station were studied to determine what portion of the variance among litters can be attributed to additive genetic differences among lines, to genetic differences among lines in their mothering abilities, to specific or "nicking" effects, and to sex linkage effects. The results of crosses were also used to obtain estimates of the general combining abilities and maternal abilities of the 12 lines and of the values of specific crosses among them. A total of 214 litters farrowed during the fall seasons of 1942-1947 inclusive furnished the data for this study. These litters represented 77 of the 132 possible reciprocal crosses and 50 of 66 possible crosses ignoring how the cross was made. The number of litters per cross varied greatly. The average inbreeding coefficient of the dams was approximately 34 per cent. The characteristics studied were litter number and litter weight at birth, 21 days, 56 days, and 154 days. In addition, estimates of the general combining abilities of the different lines with respect to individual pig weight and to livability were obtained.

Mean squares for general, maternal, specific, and sex linkage effects and for error were computed by least squares methods. These mean squares were then used to effect tests of hypotheses and to provide estimates of the variances. These latter estimates were obtained by taking expected values of the mean squares and equating the expectations to the computed mean squares. No evidence was found that sex linked genes contributed to differences in litter characteristics among single crosses. Also, there was no evidence for genetic differences among lines in their mothering abilities.
with the possible exception of 21 day litter weight. Differences among lines in their general combining abilities were small, but positive estimates of the variances were obtained for 7 of the 8 measurements. In none of the measurements did such differences account for more than 5 per cent of the variability among crosses. The estimates of variance due to general combining ability correspond to heritability estimates of 0 to 13 per cent for the original population from which the lines were formed. Since some of the lines are related, since selection has been in the same direction for all lines, and since the heritability estimates have large sampling errors, it was concluded that the estimates of general combining ability variances for litter characteristics were not materially lower than might be expected. Relationship of lines, selection in the same direction, and sampling error may also account for the failure to find maternal differences among lines. Specific effects accounted for 5 to 15 per cent of the variation among crosses in the 8 characteristics studied. These estimates seem surprisingly high in crosses among lines some of which are related and all of which come from the same breed. How much of this "nicking" effect is due to epistasis and how much to dominance effects could not be determined.

The least squares estimates of individual line and cross line effects are not the best estimates under the assumptions that the lines were randomly drawn from a population of lines. Methods were developed for utilizing least squares estimates to obtain "best" estimates for each of the lines. The method involved adjusting the least squares estimates toward 0 in accordance with the regression of the parameter on its least squares estimate. It
was shown that this is the maximum likelihood solution. Also, a method was developed for estimating the true value of an effect which is assumed to be a linear function of certain randomly drawn effects. The method involves constructing a selection index in which the independent variables are least squares estimates. This method was then applied to the prediction of the most probable value of the progeny of a particular line used as the male parent, of the progeny of a particular line used as the female parent, and of the progeny of a specific single cross. It was shown that the method when applied to selection of specific crosses in data like those of this study result in progress approximately 1.26 times as great as selection of the single cross averaging highest in a single cross test. Some further applications of the selection for specific effects to problems in animal breeding were described.

Balanced designs for testing lines in single crosses, three-way crosses, and top crosses were developed and the computational procedures described. It was concluded from an examination of these designs that these three tests are almost equally efficient for estimating general combining abilities of lines when maternal effects are real, but that cross line tests are much more efficient than the top cross test if maternal effects can be ignored, as they could be in this study. Also, the cross line tests in contrast to the top cross test furnish information concerning maternal abilities and specific and sex linkage effects. If specific effects are really as important as this study indicates, the primary emphasis in a testing program should be the testing of line crosses.

The choice of the most efficient testing program is dependent upon the following factors: the size of the testing facilities, the use to which
the lines are to be put, and the relative sizes of the various sources of variability among the progeny tested in the various types of tests. Methods were presented for choosing the best type of test for particular sets of circumstances and examples were given utilizing the variance estimates obtained in this study.

Methods for analysis of data from designs which are not balanced were described in detail as were certain computational short cuts utilized in this study. These methods may be helpful to the workers in the field of animal breeding where the data are often non-orthogonal.
VI. LITERATURE CITED


Hazel, I. N. and W. F. Lamoreux. 1947. Heritability, maternal effects
and nicking in relation to sexual maturity and body weight in

and J. L. Lush. 1948. Inbreeding hogs. Iowa Farm Sci. 2:
3-6.


giate Press, Inc.

H. O. Hetzer, and C. G. Culbertson. 1934. Factors affecting

and A. E. Molln. 1942. Litter size and weight as permanent


Sprague, G. F. and L. A. Tatum. 1942. General vs. specific combining
ability in single crosses of corn. Jour. Amer. Soc. Agron. 34:
923-932.

Tippett, L. H. C. 1925. On the extreme individuals and the range of samples


Whatley, J. A., Jr. 1937. A method for comparing the growth rate of pigs
Iowa State College Library.

Wilks, S. S. 1938. The analysis of variance and covariance in non-orthogonal
data. Metron 13: 141-158.

1944. Preliminary report on crossing of inbred lines of swine.

Wright, S. 1922. The effects of inbreeding and crossbreeding on guinea

VII. ACKNOWLEDGMENT

The author wishes to express his sincere appreciation to Dr. L. N. Hazel for his suggestion of this study and for his guidance and encouragement throughout the investigation, to Dr. J. L. Lush for the opportunity to conduct this study and for the inspiration which he has been as a teacher and adviser, to Dr. John W. Gowen for his interest and counsel, and to Professor O. Kempthorne and Dr. S. Lee Crump for suggestions regarding the problem of estimation of components of variance. The data for this study were made available by the Animal Breeding Subsection of the Iowa Agricultural Experiment Station in cooperation with the Regional Swine Breeding Laboratory of the United States Department of Agriculture.
APPENDIX I. NUMERICAL EXAMPLE OF ANALYSIS OF SINGLE CROSSES

The computational procedure for testing hypotheses, for obtaining least squares estimates of the parameters, for estimating variance components, and for adjusting least squares estimates will be illustrated by an example involving four lines and two A classes. The subclass numbers are presented in Table 38 and the subclass sums in Table 39.

Table 38

<table>
<thead>
<tr>
<th>Subclass Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line of Dam</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Line of Dam</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Sire</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Sum</td>
</tr>
</tbody>
</table>

Table 39

<table>
<thead>
<tr>
<th>Subclass Sums</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line of Dam</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Line of Dam</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Sire</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Sum</td>
</tr>
</tbody>
</table>
\[ R(T) = \sum y_{ijk}^2 = 3918. \]

A. Tests of Hypotheses

1. Test of hypothesis that \( r_{ij} = 0 \)

\[ R(\mu, g, m, s, r, a) = R(\text{reciprocal crosses, } a). \] The equations to be solved are shown in Table 40. These equations are most simply solved by first absorbing the reciprocal cross coefficients into those for \( a_k \). The computational procedure for doing this is illustrated in Table 41.

<table>
<thead>
<tr>
<th>( r_{ij} )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>3 (.60000)</td>
<td>2 (.40000)</td>
<td>42</td>
</tr>
<tr>
<td>13</td>
<td>4 (.57143)</td>
<td>3 (.42857)</td>
<td>55</td>
</tr>
<tr>
<td>21</td>
<td>2 (.40000)</td>
<td>3 (.60000)</td>
<td>36</td>
</tr>
<tr>
<td>23</td>
<td>3 (.37500)</td>
<td>5 (.62500)</td>
<td>68</td>
</tr>
<tr>
<td>24</td>
<td>5 (.55556)</td>
<td>4 (.44444)</td>
<td>58</td>
</tr>
<tr>
<td>31</td>
<td>2 (.50000)</td>
<td>2 (.50000)</td>
<td>28</td>
</tr>
<tr>
<td>32</td>
<td>4 (.57143)</td>
<td>3 (.42857)</td>
<td>45</td>
</tr>
<tr>
<td>34</td>
<td>4 (.50000)</td>
<td>4 (.50000)</td>
<td>50</td>
</tr>
<tr>
<td>41</td>
<td>4 (.57143)</td>
<td>3 (.42857)</td>
<td>41</td>
</tr>
<tr>
<td>42</td>
<td>5 (.55556)</td>
<td>4 (.44444)</td>
<td>59</td>
</tr>
<tr>
<td>43</td>
<td>2 (.50000)</td>
<td>2 (.50000)</td>
<td>24</td>
</tr>
</tbody>
</table>

The values given in Table 41 are the subclass numbers in a two-way classification of reciprocal crosses by \( A \) classes. The numbers in the parentheses are \( \frac{n_{121}}{n_{12}} \), \( \frac{n_{122}}{n_{12}} \), etc.

The \( a_1 a_1 \) coefficient shown in Table 42 is obtained as follows:

\[ 41 - 3(.60000) - 4(.57143) - \ldots - 2(.50000) = 19.062. \]
Table 40

Equations for $R(\mu, g, m, s, r, a)$

<table>
<thead>
<tr>
<th>$a_k$</th>
<th>$R_{44}$</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41 41 41</td>
<td>245</td>
</tr>
<tr>
<td>2</td>
<td>37 37 37</td>
<td>296</td>
</tr>
<tr>
<td>3</td>
<td>2 2 2 2</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>3 3 3 3</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>5 5 5 5</td>
<td>68</td>
</tr>
<tr>
<td>6</td>
<td>7 7 7 7</td>
<td>81</td>
</tr>
<tr>
<td>7</td>
<td>9 9 9 9</td>
<td>108</td>
</tr>
<tr>
<td>8</td>
<td>11 11 11</td>
<td>123</td>
</tr>
<tr>
<td>9</td>
<td>13 13 13</td>
<td>156</td>
</tr>
<tr>
<td>10</td>
<td>14 14 14</td>
<td>178</td>
</tr>
<tr>
<td>11</td>
<td>15 15 15</td>
<td>195</td>
</tr>
<tr>
<td>12</td>
<td>16 16 16</td>
<td>216</td>
</tr>
<tr>
<td>13</td>
<td>17 17 17</td>
<td>237</td>
</tr>
<tr>
<td>14</td>
<td>18 18 18</td>
<td>258</td>
</tr>
<tr>
<td>15</td>
<td>19 19 19</td>
<td>279</td>
</tr>
<tr>
<td>16</td>
<td>20 20 20</td>
<td>300</td>
</tr>
<tr>
<td>17</td>
<td>21 21 21</td>
<td>321</td>
</tr>
<tr>
<td>18</td>
<td>22 22 22</td>
<td>342</td>
</tr>
<tr>
<td>19</td>
<td>23 23 23</td>
<td>363</td>
</tr>
<tr>
<td>20</td>
<td>24 24 24</td>
<td>384</td>
</tr>
<tr>
<td>21</td>
<td>25 25 25</td>
<td>405</td>
</tr>
<tr>
<td>22</td>
<td>26 26 26</td>
<td>426</td>
</tr>
<tr>
<td>23</td>
<td>27 27 27</td>
<td>447</td>
</tr>
<tr>
<td>24</td>
<td>28 28 28</td>
<td>468</td>
</tr>
<tr>
<td>25</td>
<td>29 29 29</td>
<td>489</td>
</tr>
<tr>
<td>26</td>
<td>30 30 30</td>
<td>510</td>
</tr>
<tr>
<td>27</td>
<td>31 31 31</td>
<td>531</td>
</tr>
<tr>
<td>28</td>
<td>32 32 32</td>
<td>552</td>
</tr>
<tr>
<td>29</td>
<td>33 33 33</td>
<td>573</td>
</tr>
<tr>
<td>30</td>
<td>34 34 34</td>
<td>594</td>
</tr>
<tr>
<td>31</td>
<td>35 35 35</td>
<td>615</td>
</tr>
<tr>
<td>32</td>
<td>36 36 36</td>
<td>636</td>
</tr>
<tr>
<td>33</td>
<td>37 37 37</td>
<td>657</td>
</tr>
<tr>
<td>34</td>
<td>38 38 38</td>
<td>678</td>
</tr>
<tr>
<td>35</td>
<td>39 39 39</td>
<td>699</td>
</tr>
<tr>
<td>36</td>
<td>40 40 40</td>
<td>720</td>
</tr>
<tr>
<td>37</td>
<td>41 41 41</td>
<td>741</td>
</tr>
<tr>
<td>38</td>
<td>42 42 42</td>
<td>762</td>
</tr>
<tr>
<td>39</td>
<td>43 43 43</td>
<td>783</td>
</tr>
<tr>
<td>40</td>
<td>44 44 44</td>
<td>804</td>
</tr>
<tr>
<td>41</td>
<td>45 45 45</td>
<td>825</td>
</tr>
<tr>
<td>42</td>
<td>46 46 46</td>
<td>846</td>
</tr>
<tr>
<td>43</td>
<td>47 47 47</td>
<td>867</td>
</tr>
<tr>
<td>44</td>
<td>48 48 48</td>
<td>888</td>
</tr>
<tr>
<td>45</td>
<td>49 49 49</td>
<td>909</td>
</tr>
<tr>
<td>46</td>
<td>50 50 50</td>
<td>930</td>
</tr>
<tr>
<td>47</td>
<td>51 51 51</td>
<td>951</td>
</tr>
<tr>
<td>48</td>
<td>52 52 52</td>
<td>972</td>
</tr>
<tr>
<td>49</td>
<td>53 53 53</td>
<td>993</td>
</tr>
<tr>
<td>50</td>
<td>54 54 54</td>
<td>1014</td>
</tr>
<tr>
<td>51</td>
<td>55 55 55</td>
<td>1035</td>
</tr>
<tr>
<td>52</td>
<td>56 56 56</td>
<td>1056</td>
</tr>
<tr>
<td>53</td>
<td>57 57 57</td>
<td>1077</td>
</tr>
<tr>
<td>54</td>
<td>58 58 58</td>
<td>1098</td>
</tr>
<tr>
<td>55</td>
<td>59 59 59</td>
<td>1119</td>
</tr>
<tr>
<td>56</td>
<td>60 60 60</td>
<td>1140</td>
</tr>
<tr>
<td>57</td>
<td>61 61 61</td>
<td>1161</td>
</tr>
<tr>
<td>58</td>
<td>62 62 62</td>
<td>1182</td>
</tr>
<tr>
<td>59</td>
<td>63 63 63</td>
<td>1203</td>
</tr>
<tr>
<td>60</td>
<td>64 64 64</td>
<td>1224</td>
</tr>
<tr>
<td>61</td>
<td>65 65 65</td>
<td>1245</td>
</tr>
<tr>
<td>62</td>
<td>66 66 66</td>
<td>1266</td>
</tr>
<tr>
<td>63</td>
<td>67 67 67</td>
<td>1287</td>
</tr>
<tr>
<td>64</td>
<td>68 68 68</td>
<td>1308</td>
</tr>
<tr>
<td>65</td>
<td>69 69 69</td>
<td>1329</td>
</tr>
<tr>
<td>66</td>
<td>70 70 70</td>
<td>1350</td>
</tr>
<tr>
<td>67</td>
<td>71 71 71</td>
<td>1371</td>
</tr>
<tr>
<td>68</td>
<td>72 72 72</td>
<td>1392</td>
</tr>
<tr>
<td>69</td>
<td>73 73 73</td>
<td>1413</td>
</tr>
<tr>
<td>70</td>
<td>74 74 74</td>
<td>1434</td>
</tr>
<tr>
<td>71</td>
<td>75 75 75</td>
<td>1455</td>
</tr>
<tr>
<td>72</td>
<td>76 76 76</td>
<td>1476</td>
</tr>
<tr>
<td>73</td>
<td>77 77 77</td>
<td>1497</td>
</tr>
<tr>
<td>74</td>
<td>78 78 78</td>
<td>1518</td>
</tr>
<tr>
<td>75</td>
<td>79 79 79</td>
<td>1539</td>
</tr>
<tr>
<td>76</td>
<td>80 80 80</td>
<td>1560</td>
</tr>
<tr>
<td>77</td>
<td>81 81 81</td>
<td>1581</td>
</tr>
<tr>
<td>78</td>
<td>82 82 82</td>
<td>1602</td>
</tr>
<tr>
<td>79</td>
<td>83 83 83</td>
<td>1623</td>
</tr>
<tr>
<td>80</td>
<td>84 84 84</td>
<td>1644</td>
</tr>
<tr>
<td>81</td>
<td>85 85 85</td>
<td>1665</td>
</tr>
<tr>
<td>82</td>
<td>86 86 86</td>
<td>1686</td>
</tr>
<tr>
<td>83</td>
<td>87 87 87</td>
<td>1707</td>
</tr>
<tr>
<td>84</td>
<td>88 88 88</td>
<td>1728</td>
</tr>
<tr>
<td>85</td>
<td>89 89 89</td>
<td>1749</td>
</tr>
<tr>
<td>86</td>
<td>90 90 90</td>
<td>1770</td>
</tr>
<tr>
<td>87</td>
<td>91 91 91</td>
<td>1791</td>
</tr>
<tr>
<td>88</td>
<td>92 92 92</td>
<td>1812</td>
</tr>
<tr>
<td>89</td>
<td>93 93 93</td>
<td>1833</td>
</tr>
<tr>
<td>90</td>
<td>94 94 94</td>
<td>1854</td>
</tr>
<tr>
<td>91</td>
<td>95 95 95</td>
<td>1875</td>
</tr>
<tr>
<td>92</td>
<td>96 96 96</td>
<td>1896</td>
</tr>
<tr>
<td>93</td>
<td>97 97 97</td>
<td>1917</td>
</tr>
<tr>
<td>94</td>
<td>98 98 98</td>
<td>1938</td>
</tr>
<tr>
<td>95</td>
<td>99 99 99</td>
<td>1959</td>
</tr>
<tr>
<td>96</td>
<td>100 100 100</td>
<td>1980</td>
</tr>
</tbody>
</table>
The \( a_1 a_2 \) coefficient = \( -3(-.40000) - 4(.42857) - \ldots - 2(.50000) \)
\[ = -19.062. \]

The \( a_2 a_2 \) coefficient = \( 37 - 2(-.40000) - 3(.42857) - \ldots - 2(.50000) \)
\[ = 19.062. \]

The sum associated with \( a_1 = 245 - 42(.60000) - 55(.57143) - \ldots - 24(.50000) = -37.672. \]

The \( a_2 \) sum = \( 296 - 42(.40000) - 55(.42857) - \ldots - 24(.50000) = 37.672. \)

<p>| Table 42 |
|---|---|---|---|
| Equations for ( R(\mu, g, m, s, r, a) ) with ( \hat{\gamma}_{ij} ) absorbed |</p>
<table>
<thead>
<tr>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>19.062</td>
<td>-19.062</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>-19.062</td>
<td>19.062</td>
</tr>
</tbody>
</table>

Since \( a_1 + a_2 = 0 \), the \( a_2 \) equation can be deleted and the \( a_2 \) coefficient can be subtracted from the \( a_1 \) coefficient in the \( a_1 \) equation, giving
\[ 38.124 a_1 = -37.672. \]

Therefore, \( a_1 = -.9881 \)
\[ a_2 = .9881 \]

Substituting these values for \( a_k \) in the equations of Table 40 \( \hat{\gamma}_{12} = \frac{1}{5} \)
\[ 42 - 3(-.9881) - 2(.9881) = 8.5976. \]

The complete set of \( \hat{\gamma}_{ij} \) estimates is
\[ 12 = 8.5976 \quad 23 = 6.5542 \quad 34 = 6.2500 \]
\[ 13 = 7.9983 \quad 24 = 7.1976 \quad 41 = 5.9983 \]
\[ 14 = 7.0024 \quad 31 = 7.0000 \quad 42 = 6.6653 \]
\[ 21 = 8.2530 \quad 32 = 6.5697 \quad 43 = 6.0000 \]
The reduction due to fitting these parameters = 245(-.9881) + 296 (.9881) + 42(.5976) + ..... + 24(6.0000) = 3884.07.

\[ R(\mu, g, m, s, a) = R(\text{crosses}, m, a). \]  
Table 43 shows the equations for estimating these parameters.

**Table 43**

Least Squares Equations for \( R(\mu, g, m, s, a) \)

<table>
<thead>
<tr>
<th>( i )</th>
<th>( a_{1j} )</th>
<th>( a_{12} )</th>
<th>( a_{13} )</th>
<th>( a_{14} )</th>
<th>( a_{22} )</th>
<th>( a_{23} )</th>
<th>( a_{24} )</th>
<th>( a_{32} )</th>
<th>( a_{33} )</th>
<th>( a_{34} )</th>
<th>( a_{42} )</th>
<th>( a_{43} )</th>
<th>( a_{44} )</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>23</td>
<td>24</td>
<td>34</td>
<td>( a_1 )</td>
<td>( a_2 )</td>
<td>( m_1 )</td>
<td>( m_2 )</td>
<td>( m_3 )</td>
<td>( m_4 )</td>
<td>( \text{Sum} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>23</td>
<td>24</td>
<td>34</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>12</td>
<td>16</td>
<td>14</td>
<td>12</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td>16</td>
<td>14</td>
<td>12</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this set of equations it is advisable to first absorb the \( a_{ij} \) coefficients since they are the most numerous. Table 44 illustrates the first step in the absorption.

**Table 44**

Computational Procedure for Absorbing \( a_{ij} \)

<table>
<thead>
<tr>
<th>( a_{ij} )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_3 )</th>
<th>( m_4 )</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>6 (.46154)</td>
<td>7 (.53846)</td>
<td>8 (.61538)</td>
<td>5 (.38462)</td>
<td>( \text{Sum} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>6 (.54545)</td>
<td>5 (.45455)</td>
<td>4 (.36364)</td>
<td>7 (.63636)</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>6 (.50000)</td>
<td>6 (.50000)</td>
<td>7 (.58333)</td>
<td>5 (.41667)</td>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>9 (.56250)</td>
<td>7 (.43750)</td>
<td>7 (.43750)</td>
<td>9 (.56250)</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>8 (.57143)</td>
<td>6 (.42857)</td>
<td>9 (.64286)</td>
<td>5 (.35714)</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>6 (.50000)</td>
<td>6 (.50000)</td>
<td>4 (.33333)</td>
<td>8 (.66667)</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The coefficient of $a_1 a_1$ is $41 - 6(0.46154) - \cdots - 6(0.50000) = 19.324$.

The coefficient of $a_1 a_2$ is $-6(0.53846) - \cdots - 6(0.50000) = -19.324$.

The coefficient of $a_1 m_1$ is $9 - 6(0.61538) - 6(36364) - 6(0.58333) = -0.374$.

The sum associated with $a_1$ is $245 - 110(0.46154) - \cdots - 74 (0.50000) = -38.194$.

The complete set of equations resulting from the absorption of the $s_{ij}$ coefficients is shown in Table 45.

Table 45

<table>
<thead>
<tr>
<th>$s_{ij}$ absorbed</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$m_3$</th>
<th>$m_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>19.324</td>
<td>-19.324</td>
<td>-0.374</td>
<td>0.612</td>
<td>0.119</td>
<td>-0.357</td>
</tr>
<tr>
<td>$a_2$</td>
<td>-19.324</td>
<td>19.324</td>
<td>0.374</td>
<td>-0.612</td>
<td>-0.119</td>
<td>0.357</td>
</tr>
<tr>
<td>$m_1$</td>
<td>-0.374</td>
<td>0.374</td>
<td>8.539</td>
<td>-3.077</td>
<td>-2.545</td>
<td>-2.917</td>
</tr>
<tr>
<td>$m_2$</td>
<td>0.612</td>
<td>-0.612</td>
<td>-3.077</td>
<td>10.229</td>
<td>-3.938</td>
<td>-3.214</td>
</tr>
<tr>
<td>$m_3$</td>
<td>0.119</td>
<td>-0.119</td>
<td>-2.545</td>
<td>-3.938</td>
<td>9.150</td>
<td>-2.667</td>
</tr>
<tr>
<td>$m_4$</td>
<td>-0.357</td>
<td>0.357</td>
<td>-2.917</td>
<td>-3.214</td>
<td>-2.667</td>
<td>8.798</td>
</tr>
</tbody>
</table>

It will be noted that the coefficients for a particular set of parameters add to zero in each equation, also that the sums associated with a particular set of parameters add to zero. This fact serves as a useful check on the accuracy of the computations.

Next the $a_2$ and $m_4$ equations are deleted, and in the remaining equations the coefficients of $a_2$ are subtracted from the $a_1$ coefficients and
the m₄ coefficients from the coefficients of m₁, m₂, and m₃. When this is done, the equations of Table 46 result.

The iterative method was used to solve the equations. The successive estimates arising in the solution are shown just above the equations. The first estimate used for each parameter was zero.

Table 46

Reduced Equations for R(μ, g, m, s, a)
Including Steps in the Iterative Solution

<table>
<thead>
<tr>
<th>Successive Estimates</th>
<th>5</th>
<th>- .9897</th>
<th>-.5755</th>
<th>-.0295</th>
<th>.1562</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>- .9897</td>
<td>-.5755</td>
<td>-.0295</td>
<td>.1562</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>- .9895</td>
<td>-.5756</td>
<td>-.0296</td>
<td>.1562</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>- .9883</td>
<td>-.5699</td>
<td>-.0381</td>
<td>.1552</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>a₁</th>
<th>m₁</th>
<th>m₂</th>
<th>m₃</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>38.648</td>
<td>-.017</td>
<td>.699</td>
<td>.476</td>
<td>-38.194</td>
</tr>
<tr>
<td>m₁</td>
<td>-.748</td>
<td>11.456</td>
<td>-.160</td>
<td>.372</td>
<td>-5.790</td>
</tr>
<tr>
<td>m₂</td>
<td>1.224</td>
<td>.137</td>
<td>13.443</td>
<td>-.724</td>
<td>-1.800</td>
</tr>
<tr>
<td>m₃</td>
<td>.238</td>
<td>.122</td>
<td>-.1271</td>
<td>11.817</td>
<td>1.578</td>
</tr>
</tbody>
</table>

Therefore, the least squares estimates of the a's and m's are as follows:

a₁ = -.9897,   m₁ = -.5755,   m₃ = .1562
a₂ = .9897,   m₂ = -.0295,   m₄ = .4488

Substituting these estimates in the  wid hat subscripts i j  equations of Table 43, the estimate of  wid hat subscripts i j  ₁₂ = \frac{1}{13} \left[ 110 - 6(-.9897) - 7(.9897) - 8(-.5755) - 5(-.0295) \right] = 8.7509.

The other  wid hat subscripts i j  are:

13 = 7.7453,  23 = 6.4862,  34 = 5.8154
14 = 6.5654,  24 = 6.7144
The test of significance of the differences among reciprocal crosses corrected for the maternal effect can now be made. The sum of squares among \( R = R(\mu, g, m, s, r, a) - R(\mu, g, m, s, a) \) = 3884.07 - 3883.61 = 0.46. The error sum of squares is 

\[
\sum y_{ijk}^2 - R(\text{reciprocal crosses, a}) = 3918 - 3884.07 = 33.93.
\]

The degrees of freedom for \( R = 13 - 10 = 3 \). The 13 comes from the number of independent parameters in \( R(\text{reciprocal crosses, a}) \). They include \( a_1 \) and 12 for reciprocal crosses. The 10 comes from the number of independent parameters in \( R(\text{crosses, m, a}) \). They are \( a_1, m_1, m_2, m_3, \) and 6 for the \( y_{ij} \). The degrees of freedom for error = 78 - 13 = 65. 78 is the total number of observations.

The analysis of variance is as follows:

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among R</td>
<td>3</td>
<td>0.46</td>
<td>0.153</td>
</tr>
<tr>
<td>Error</td>
<td>65</td>
<td>33.93</td>
<td>0.522</td>
</tr>
</tbody>
</table>

Consequently, the null hypothesis, \( r_{ij} = 0 \), is accepted.

2. Test of hypothesis that \( a_{ij} = 0 \) assuming \( r_{ij} = 0 \)

The hypothesis that \( a_{ij} = 0 \) is now tested. This test requires \( R(\mu, g, m, s, a) - R(\mu, g, m, a) \). Let \( d_j = \mu + g_j + m_j \). Then the equations needed for \( R(\mu, g, m, a) \) are the ones shown in Table 47. The \( d_j \) coefficients can now be absorbed in the manner illustrated for absorbing \( k_{ij} \) and \( s_{ij} \). The solution to the resulting set of equations is

\[
\hat{a}_1 = -0.9949 \\
\hat{a}_2 = 0.9949 \\
\hat{g}_1 = 1.0644 \\
\hat{g}_2 = 0.4352 \\
\hat{g}_3 = -0.5344 \\
\hat{g}_4 = -0.9652
\]
Substituting these values in the equations of Table 47 gives the following values for the \( \hat{d}_j \).

\[
\begin{align*}
1 & = 7.4431 \\
2 & = 7.4329 \\
3 & = 6.5741 \\
4 & = 6.5432
\end{align*}
\]

Hence, \( R(\mu, g, m, a) = 245(-.9946) + 296(.9946) + \ldots + 121(6.5432) = 3874.24 \), and the sum of squares for \( S = 3883.61 - 3874.24 = 9.37 \). The sum of squares for error = 3918 - 3883.61 = 34.39.

The analysis of variance is as follows:

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among S</td>
<td>2</td>
<td>9.37</td>
<td>4.685</td>
</tr>
<tr>
<td>Error</td>
<td>68</td>
<td>34.39</td>
<td>0.506</td>
</tr>
</tbody>
</table>

The hypothesis that \( a_{ij} = 0 \) is rejected.

Table 47

<table>
<thead>
<tr>
<th></th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( g_1 )</th>
<th>( g_2 )</th>
<th>( g_3 )</th>
<th>( g_4 )</th>
<th>( d_1 )</th>
<th>( d_2 )</th>
<th>( d_3 )</th>
<th>( d_4 )</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>41</td>
<td>0</td>
<td>9</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>9</td>
<td>12</td>
<td>11</td>
<td>9</td>
<td>245</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>0</td>
<td>37</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>296</td>
</tr>
<tr>
<td>( g_1 )</td>
<td>9</td>
<td>8</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>133</td>
</tr>
<tr>
<td>( g_2 )</td>
<td>11</td>
<td>11</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>9</td>
<td>5</td>
<td>161</td>
</tr>
<tr>
<td>( g_3 )</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td>6</td>
<td>123</td>
</tr>
<tr>
<td>( g_4 )</td>
<td>11</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>7</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>124</td>
</tr>
<tr>
<td>( d_1 )</td>
<td>1</td>
<td>9</td>
<td>10</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>7</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>137</td>
</tr>
<tr>
<td>( d_2 )</td>
<td>12</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>9</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>146</td>
</tr>
<tr>
<td>( d_3 )</td>
<td>11</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>137</td>
</tr>
<tr>
<td>( d_4 )</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>121</td>
</tr>
</tbody>
</table>

3. Test of hypothesis that \( a_{ij} = 0 \) assuming \( m_j = r_{ij} = 0 \)

In many cases, such as with plants and perhaps with poultry, \( m_j = 0 \) may be so logical an hypothesis that it does not require testing. In other cases it may be tested and accepted. In these cases the sum of squares for

\[ S = R(\mu, g, s, a) - R(\mu, g, a). \]
The equations needed for the first of these reductions is obtained by deleting the m equations and m unknowns from the equations of Table 43.

The solution is

\[ \hat{a}_1 = -0.9883 \]
\[ \hat{a}_2 = 0.9883 \]
\[ \hat{a}_{12} = 8.3855 \]
\[ \hat{a}_{13} = 7.6353 \]

\[ R(\mu, g, s, a) = 245(-0.9883) + \ldots + 74(6.1667) = 3877.16. \]

The equations for \( R(\mu, g, s) \) are presented in Table 48.

**Table 48**

<table>
<thead>
<tr>
<th>Equations for ( R(\mu, g, s) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu + a_1 )</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>( \mu + a_1 )</td>
</tr>
<tr>
<td>( \mu + a_2 )</td>
</tr>
<tr>
<td>( g_1 )</td>
</tr>
<tr>
<td>( g_2 )</td>
</tr>
<tr>
<td>( g_3 )</td>
</tr>
<tr>
<td>( g_4 )</td>
</tr>
</tbody>
</table>

It does not appear possible to absorb the \( g_1 \) coefficients, and the \( a_k \) are too few to be worth absorbing. The solution to the equations is

\[ \hat{a}_1 = 5.9930 \]
\[ \hat{a}_2 = -3.818 \]
\[ \hat{a}_3 = -3.804 \]
\[ \hat{a}_4 = -7.403 \]

Hence, \( R(\mu, g, a) = 245(5.9930) + \ldots + 245(-7.403) = 3866.63. \)

The sum of squares for \( S = 3877.16 - 3866.63 = 10.53 \), with \( 7 - 5 = 2 \) degrees of freedom.
The error sum of squares = $3918 - 3877.16 = 40.84$, with $78 - 7 = 71$ degrees of freedom. The analysis of variance is as follows:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among S</td>
<td>2</td>
<td>10.53</td>
<td>5.265</td>
</tr>
<tr>
<td>Error</td>
<td>71</td>
<td>40.84</td>
<td>.575</td>
</tr>
</tbody>
</table>

The hypothesis that $\varepsilon_{ij} = 0$ is rejected.

4. Test of hypothesis that $m_j = 0$ assuming $\varepsilon_{ij} = 0$

The sum of squares appropriate to this test is $R(\mu, g, m, s, a) - R(\mu, g, s, a)$. These two reductions have already been computed. Consequently, the M sum of squares = $3883.61 - 3877.16 = 6.45$, with 3 degrees of freedom.

The error sum of squares is $R(T) - R(\mu, g, m, s, a) = 3918 - 3883.61 = 34.39$ with $78 - 10 = 68$ degrees of freedom. The analysis of variance is

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among M</td>
<td>3</td>
<td>6.45</td>
<td>2.150</td>
</tr>
<tr>
<td>Error</td>
<td>68</td>
<td>34.39</td>
<td>.506</td>
</tr>
</tbody>
</table>

The hypothesis that $m_j = 0$ is rejected.

5. Test of hypothesis that $m_j = 0$ assuming $\varepsilon_{ij} = r_{ij} = 0$

If it had appeared that $\varepsilon_{ij} = 0$, a better test of the hypothesis that $m_j = 0$ than the one described above is $R(\mu, g, m, a) - R(\mu, g, a)$. Each of these $R(\ )$'s has been computed. Using these values the sum of squares among $M$ under this assumption is $3874.24 - 3866.63 = 7.61$.

The error sum of squares is $3918 - 3874.24 = 43.76$ with $78 - 8 = 70$ degrees of freedom. The analysis of variance is

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among M</td>
<td>3</td>
<td>7.61</td>
<td>2.537</td>
</tr>
<tr>
<td>Error</td>
<td>70</td>
<td>43.76</td>
<td>.625</td>
</tr>
</tbody>
</table>
6. Test of hypothesis that $g_1 = 0$ assuming $s_{ij} = r_{ij} = 0$

The tests of hypotheses that $g_1 = 0$ under the assumptions that $s_{ij} \neq 0$ and that $s_{ij}, r_{ij} \neq 0$ require an excessive amount of labor. Consequently, tests of $g_1$ illustrated in this Appendix are effected under the assumption that $s_{ij} = r_{ij} = 0$. The appropriate sum of squares for this test is $R(\mu, g, m, a) - R(\mu, m, a)$. $R(\mu, m, a)$ is computed from the equations of Table 47 from which the $g_1$ have been deleted. Its value is 3834.11. Consequently, the sum of squares for $g$ is $3874.24 - 3834.11 = 40.13$. The error sum of squares is $3918 - 3874.24 = 43.76$, with 78 - 8 = 70 degrees of freedom.

The analysis of variance then is as follows:

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among G</td>
<td>3</td>
<td>40.13</td>
<td>13.377</td>
</tr>
<tr>
<td>Error</td>
<td>70</td>
<td>43.76</td>
<td>.625</td>
</tr>
</tbody>
</table>

7. Test of hypothesis that $g_1 = 0$ assuming $s_{ij} = r_{ij} = m_j = 0$

The sum of squares for $G = R(\mu, g, a) - R(\mu, a)$. $R(\mu, a) = \frac{1}{41} (245)^2 + \frac{1}{37} (296)^2 = 3832.02$. Therefore, sum of squares for $G = 3866.63 - 3832.02 = 34.61$. The error sum of squares = $3918 - 3866.63 = 51.37$ with 78 - 5 = 73 degrees of freedom. The analysis of variance is as follows:

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among G</td>
<td>3</td>
<td>34.61</td>
<td>11.537</td>
</tr>
<tr>
<td>Error</td>
<td>73</td>
<td>51.37</td>
<td>.704</td>
</tr>
</tbody>
</table>

B. Least Squares Estimates

1. Estimates of $k_1, m_j, s_{ij}, r_{ij}$

Although in this example the hypothesis that $r_{ij} = 0$ was accepted, the method for obtaining estimates when $r_{ij} \neq 0$ will be described since such estimates might be needed in some problems. In section A 1 of this Appendix least squares estimates were computed for reciprocal crosses: $\hat{r}_{ij} = \mu + g_1 +$
Using these estimates the equations of Table 49 are set up from which estimates of \( m_j \) and \( \hat{a}_{ij} = \mu + g_i + g_j + a_{ij} \) are computed.

The \( m_1 \) equation is obtained by adding \( r_{21}, r_{31}, r_{41} \), the estimates containing \( m_1 \). The sum of these three least squares estimates also includes \( a_{12}, a_{13}, \) and \( a_{14} \), and the numerical value of this sum is 21.2513.

The \( \hat{a}_{12} \) equation comes from adding \( r_{12} \) and \( r_{21} \). These two together contain \( m_1, m_2, \) and 2 \( a_{12} \). Their sum is 16.8506.

The \( r_{ij} \) drop out of the equations because \( \sum r_{ij} = 0 \) for all \( j \) and \( r_{ij} + r_{ji} = 0 \) for all \( i \) and \( j \).

<table>
<thead>
<tr>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_3 )</th>
<th>( m_4 )</th>
<th>( \hat{a}_{12} )</th>
<th>( \hat{a}_{13} )</th>
<th>( \hat{a}_{14} )</th>
<th>( \hat{a}_{23} )</th>
<th>( \hat{a}_{24} )</th>
<th>( \hat{a}_{34} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \hat{m}_1 ) = -.5868</td>
<td>( \hat{a}_{12} ) = 8.7402</td>
<td>( \hat{a}_{23} ) = 6.4918</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{m}_2 ) = -.0430</td>
<td>( \hat{a}_{13} ) = 7.7010</td>
<td>( \hat{a}_{24} ) = 6.7296</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{m}_3 ) = .1832</td>
<td>( \hat{a}_{14} ) = 6.5704</td>
<td>( \hat{a}_{34} ) = 5.8101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{m}_4 ) = .4466</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These equations can easily be solved by absorbing the \( a_{ij} \). The solution to the set is as follows:
The estimates of $\mu$ and $g_1$ now come from the $\hat{s}_{ij}$. The equations for these estimates are shown in Table 50. The $\mu$ equation arises from the summation of all the $\hat{s}_{ij}$. This sum = 42.0431 and has in it 6 $\mu$ and 3 of each of the $g_1$. The $g_1$ equation comes from the summation of all $\hat{s}_{ij}$ containing $g_1$, and similarly for the other $g_1$ equations.

Table 50

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
<th>$g_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$g_1$</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$g_2$</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$g_3$</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$g_4$</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42.0431</td>
<td>23.0116</td>
<td>21.9616</td>
<td>20.0029</td>
</tr>
</tbody>
</table>

Since $\sum \hat{s}_{ij} = 0$, the equations of Table 50 reduce to those shown in Table 51.

Table 51

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$g_1$</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$g_2$</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>$g_3$</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>42.0431</td>
<td>23.0116</td>
<td>21.9616</td>
<td>20.0029</td>
</tr>
</tbody>
</table>

The solution to these equations is

\[ \hat{\mu} = 7.0072 \]
\[ \hat{g}_1 = 0.9950 \]
\[ \hat{g}_2 = 0.4700 \]
\[ \hat{g}_3 = -0.5094 \]
\[ \hat{g}_4 = -0.9556 \]
Then \( \hat{s}_{12} = \hat{s}_{12} - \hat{\mu} - \hat{e}_1 - \hat{e}_2 = 8.7402 - 7.0072 - .9950 - .4700 = .2680 \).

Similarly, the other \( \hat{s}_{ij} \) are

- \( \hat{s}_{13} = .2082 \)
- \( \hat{s}_{14} = -.4762 \)
- \( \hat{s}_{23} = -.4760 \)

\( \hat{r}_{12} = \hat{r}_{12} - \hat{m}_2 - \hat{e}_{12} = 5.5976 - (-.0430) - 8.7402 = -.0996 \).

Similarly, the other \( \hat{r}_{ij} \) are

- \( \hat{r}_{13} = .1141 \)
- \( \hat{r}_{14} = -.0146 \)
- \( \hat{r}_{21} = .0996 \)
- \( \hat{r}_{23} = -.1208 \)
- \( \hat{r}_{24} = .0214 \)
- \( \hat{r}_{31} = -.1142 \)

It will be noted that the \( \hat{e}_{ij} \) sum to zero over each line, that the \( \hat{r}_{ij} \) sum to zero over each line of dam and over each line of boar, and that \( \hat{r}_{ij} + \hat{r}_{ji} = 0 \). These restrictions were, of course, imposed on the estimates.

2. Estimates of \( g_{ij}, m_j, e_{ij} \)—assuming \( r_{ij} = 0 \)

In the computation of \( R(\mu, g, m, s, a) \) estimates arose of \( m_j \) and \( \hat{e}_{ij} \)\( = \mu + g_i + e_j + s_{ij} \). The values obtained for \( \hat{m}_j \) were

- \( \hat{m}_1 = -.5755 \)
- \( \hat{m}_2 = -.0295 \)
- \( \hat{m}_3 = .1562 \)
- \( \hat{m}_4 = .4488 \)

The \( \hat{e}_{ij} \) were:

- \( \hat{e}_{12} = 8.7509 \)
- \( \hat{e}_{13} = 7.7453 \)
- \( \hat{e}_{14} = 6.5654 \)
- \( \hat{e}_{24} = 6.7144 \)
- \( \hat{e}_{23} = 6.4862 \)
- \( \hat{e}_{34} = 5.8154 \)
\[ \mu \text{ and } \hat{g}_1 \text{ are estimated from the } \hat{g}_{ij}. \text{ The left members of the required equations are the same as in Table 50. The right members for each of the leading equations are:} \]

\[ \begin{align*}
\mu &: 8.7509 + \ldots + 5.8154 \\
\hat{g}_1 &: 8.7509 + 7.7453 + 6.5654 \\
\hat{g}_2 &: 8.7509 + 6.4862 + 6.7144 \\
\hat{g}_3 &: 7.7453 + 6.4862 + 5.8154 \\
\hat{g}_4 &: 6.5654 + 6.7144 + 5.8154 \\
\end{align*} \]

The solution is:

\[ \begin{align*}
\hat{\mu} &= 7.0129 & \hat{\hat{g}}_3 &= -0.4959 \\
\hat{\hat{g}}_1 &= 1.0114 & \hat{\hat{g}}_4 &= -0.9719 \\
\hat{\hat{g}}_2 &= 0.4564 &
\end{align*} \]

3. *Estimates of \( \mu, \hat{g}_1, m_j \) assuming \( \hat{g}_{ij} = \hat{r}_{ij} = 0 \)

In the computation of \( R(\mu, \hat{g}, m, a) \) estimates were obtained of \( \hat{g}_1 \) and \( \hat{d}_j \) where \( \hat{d}_j = \mu + \hat{g}_j + m_j \). Since \( \hat{\hat{g}}_1 = \hat{\epsilon}_j = 0, \hat{\hat{d}}_j = \hat{r}_j = \hat{d}_j, \) and \( \hat{\mu} = \frac{1}{p} \sum \hat{\hat{d}}_j \).

In the present example, therefore, \( \hat{\mu} = \frac{1}{7} (7.4431 + \ldots + 6.5432) = 6.9983 \).

Then since \( m_j = \hat{d}_j - \mu - \hat{g}_j \),

\[ \hat{m}_1 = 7.4431 - 6.9983 - 1.0644 = -0.6196 \]

Similarly,

\[ \begin{align*}
\hat{m}_2 &= -0.0006 \\
\hat{m}_3 &= 0.1102 \\
\hat{m}_4 &= 0.5101 \\
\end{align*} \]
C. Sampling Errors

Sampling errors will be computed under the assumption that \( r_{ij} = 0 \). What is needed is the inverse matrix of the matrix of coefficients for \( R(\mu, \epsilon, m, s, a) \). The appropriate least squares equations were presented in Table 43. The coefficients of \( \hat{e}_{ij} \) were absorbed to obtain the equations of Table 45. Then in each of the equations of Table 45 the coefficient of \( a_2 \) is subtracted from that of \( a_1 \) and the coefficient of \( m_4 \) is subtracted from those of \( m_1, m_2, \) and \( m_3 \). The \( a_2 \) and \( m_4 \) equations are deleted. The resulting set of coefficients is equated to an appropriate matrix as shown in Table 52.

Table 52

<table>
<thead>
<tr>
<th>( i )</th>
<th>( a_1 )</th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_3 )</th>
<th>( a_1 )</th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38.648</td>
<td>-.017</td>
<td>.969</td>
<td>.476</td>
<td>.50</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>m_1</td>
<td>-.748</td>
<td>11.496</td>
<td>-.160</td>
<td>.372</td>
<td>0</td>
<td>.75</td>
<td>-.25</td>
<td>-.25</td>
</tr>
<tr>
<td>m_2</td>
<td>1.224</td>
<td>.137</td>
<td>13.443</td>
<td>-.724</td>
<td>0</td>
<td>-.25</td>
<td>.75</td>
<td>-.25</td>
</tr>
<tr>
<td>m_3</td>
<td>.238</td>
<td>.122</td>
<td>-1.271</td>
<td>11.617</td>
<td>0</td>
<td>-.25</td>
<td>-.245</td>
<td>.75</td>
</tr>
</tbody>
</table>

The solution to the first set of equations is the first row and column of the invert matrix, the solution to the second set is the second row and column of the invert matrix, etc. These equations can be solved very quickly by the iterative method. For example, when the left members are equated to \( \begin{pmatrix} .50 \\ 0 \\ 0 \\ 0 \end{pmatrix} \), the successive estimates are
The complete solution is the invert matrix with respect to $a_i$ and $m_j$.

Its value is as follows:

<table>
<thead>
<tr>
<th>$i$</th>
<th>$a_1$</th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$m_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.01294</td>
<td>0.00084</td>
<td>-0.00119</td>
<td>-0.00040</td>
</tr>
<tr>
<td>3</td>
<td>0.01297</td>
<td>0.00084</td>
<td>-0.00121</td>
<td>-0.00040</td>
</tr>
<tr>
<td>4</td>
<td>0.01297</td>
<td>0.00084</td>
<td>-0.00121</td>
<td>-0.00040</td>
</tr>
</tbody>
</table>

Since $V(\hat{a}_i) + CV(\hat{a}_i \hat{a}_2) = V(\hat{m}_1) + CV(\hat{m}_1 \hat{m}_2) + CV(\hat{m}_1 \hat{m}_3) + CV(\hat{m}_1 \hat{m}_4) = 0$ and similarly with respect to other sums of variances and covariances, the elements of the $a_i$ and $m_4$ rows and columns can be computed. For example, the inverse element $a_i a_4 = -0.00084 - 0.00121 - 0.00040 = -0.00077$. The remaining elements of the $a_k m_j$ block are shown in Table 53. The elements pertaining to $\hat{a}_{ij}$ also presented in Table 53 are computed as described in Section II C 4. For example, $a_{12} a_1 = -\frac{1}{13} \left[ 6(0.01297) + 7(-0.01297) + 8(0.00084) + 5(-0.00121) \right] = 0.00095$. After the $a_{ij} a_k$ elements have been computed, $a_{ij} a_{ij}$ can be found. For example, $a_{12} a_{12} = \frac{1}{13} \left[ 1 - 6(0.00095) - 7(-0.00095) - 8(-0.03262) - 5(-0.00886) \right] = 0.10040$. If now we wish the sampling errors of $\hat{\mu}$, $\hat{e}_i$, and $\hat{a}_{ij}$ rather than of $\hat{a}_{ij}$; $\hat{\mu}$, $\hat{e}_i$, and $\hat{a}_{ij}$ need to be expressed as linear functions of $\hat{a}_{ij}$. It can be seen from an inspection of
Table 53

Inverse Matrix for $R(\mu, g, m, s, a)$
(multiplied by 100,000)

<table>
<thead>
<tr>
<th></th>
<th>$s_{12}$</th>
<th>$s_{13}$</th>
<th>$s_{14}$</th>
<th>$s_{23}$</th>
<th>$s_{24}$</th>
<th>$s_{34}$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$m_3$</th>
<th>$m_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{12}$</td>
<td>10,040</td>
<td>-130</td>
<td>1,039</td>
<td>-788</td>
<td>-198</td>
<td>-2,067</td>
<td>95</td>
<td>-95</td>
<td>-3,262</td>
<td>-866</td>
<td>2,054</td>
<td>2,074</td>
</tr>
<tr>
<td>$s_{13}$</td>
<td>11,377</td>
<td>-421</td>
<td>955</td>
<td>-1,883</td>
<td>-466</td>
<td>-123</td>
<td>123</td>
<td>-876</td>
<td>1,714</td>
<td>-3,074</td>
<td>2,236</td>
<td></td>
</tr>
<tr>
<td>$s_{14}$</td>
<td>10,656</td>
<td>-2,206</td>
<td>-780</td>
<td>162</td>
<td>-81</td>
<td>81</td>
<td>-2,963</td>
<td>2,024</td>
<td>2,365</td>
<td>-1,426</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{23}$</td>
<td>8,538</td>
<td>249</td>
<td>-489</td>
<td>-87</td>
<td>87</td>
<td>2,246</td>
<td>-1,562</td>
<td>-2,834</td>
<td>2,150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{24}$</td>
<td>9,353</td>
<td>-1,35</td>
<td>135</td>
<td>2,076</td>
<td>-2,834</td>
<td>1,792</td>
<td>-1,034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{34}$</td>
<td>10,845</td>
<td>-38</td>
<td>38</td>
<td>2,222</td>
<td>1,812</td>
<td>-532</td>
<td>-3,502</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{1}$</td>
<td>1,297</td>
<td>-1,297</td>
<td>84</td>
<td>-121</td>
<td>-40</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{2}$</td>
<td>1,297</td>
<td>-84</td>
<td>121</td>
<td>40</td>
<td>-77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{1}$</td>
<td>6,602</td>
<td>-2,064</td>
<td>-2,407</td>
<td>-2,131</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{2}$</td>
<td>5,531</td>
<td>-1,497</td>
<td>-1,970</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{3}$</td>
<td>6,211</td>
<td>-2,307</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{4}$</td>
<td>6,408</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the equations described in Section B 2 of this appendix that the linear functions presented in Table 54 are the proper ones for computing the least squares estimates of \( \mu, g_1, \) and \( s_{ij}. \)

Table 54

Parameter Estimates Expressed as Linear Functions of \( \hat{s}_{ij} \)

<table>
<thead>
<tr>
<th>Least Squares Estimate</th>
<th>( \hat{s}_{12} )</th>
<th>( \hat{s}_{13} )</th>
<th>( \hat{s}_{14} )</th>
<th>( \hat{s}_{23} )</th>
<th>( \hat{s}_{24} )</th>
<th>( \hat{s}_{34} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
<tr>
<td>( g_1 )</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>-1/4</td>
<td>-1/4</td>
<td>-1/4</td>
</tr>
<tr>
<td>( g_2 )</td>
<td>1/4</td>
<td>-1/4</td>
<td>-1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>-1/4</td>
</tr>
<tr>
<td>( g_3 )</td>
<td>-1/4</td>
<td>1/4</td>
<td>-1/4</td>
<td>1/4</td>
<td>-1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>( g_4 )</td>
<td>-1/4</td>
<td>-1/4</td>
<td>1/4</td>
<td>-1/4</td>
<td>1/4</td>
<td>-1/4</td>
</tr>
<tr>
<td>( a_{12} )</td>
<td>1/3</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
</tr>
<tr>
<td>( a_{13} )</td>
<td>-1/6</td>
<td>1/3</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
</tr>
<tr>
<td>( a_{14} )</td>
<td>-1/6</td>
<td>-1/6</td>
<td>1/3</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
</tr>
<tr>
<td>( a_{23} )</td>
<td>-1/6</td>
<td>1/3</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
</tr>
<tr>
<td>( a_{24} )</td>
<td>1/3</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
</tr>
<tr>
<td>( a_{34} )</td>
<td>-1/6</td>
<td>-1/6</td>
<td>1/3</td>
<td>-1/6</td>
<td>-1/6</td>
<td>-1/6</td>
</tr>
</tbody>
</table>

Now the inverse elements with respect to \( g_1 \) can be computed from the information available in Tables 53 and 54. The method for doing this is analogous to the method of finding the variances and covariances of linear functions. For example, \( \mu^{-1} \frac{1}{6^2} \left( \frac{1}{10,040} \right) + 2(-130) + \ldots + 2(92) = .01306. \)

D. Variance Components

Since the tests of significance resulted in accepting the hypothesis that \( r_{ij} = 0 \) and rejecting the hypotheses that the other sets of parameters \( = 0, \) the estimates of \( \sigma_g^2, \sigma_m^2, \sigma_s^2, \) and \( \sigma_o^2 \) will be effected under the assumption that \( \sigma_r^2 = 0. \)
1. Estimation of $\sigma^2$

$$\hat{\sigma}^2 = \left[ R(T) - R(\mu, g, m, s, a) \right] / \text{d.f. for error} = (3918 - 3883.61) / 68 = .506.$$ 

2. Estimation of $\sigma^2$

$$\hat{\sigma}^2 = (1/k_1) \left[ R(\mu, g, m, s, a) - R(\mu, g, m, a) - k_2 \sigma^2 \right].$$

$k_1 = n \ldots$ - coefficient of $\sigma^2$ in $E R(\mu, g, m, a)$.

$k_2 = \text{d.f. for S sum of squares.}$

The computation of $k_1$ can be done in a systematic way. Table 55 is first prepared. This table shows the values of the coefficients of the $s_{ij}$ in the sums associated with $a_k, g_1,$ and $d_j$. These associated sums are the right members of the equations of Table 47. Let us consider the expectation of $y_{..1}$, the sum associated with $a_1$. Now this sum includes 3 observations from the cross of line 1 males by line 2 females and 3 observations from the reciprocal of this cross. Consequently, the coefficient of $s_{12}$ in the expected value of the sum associated with $a_1$ is $3 + 3 = 6$. The other entries in Table 55 are determined in a similar fashion.

### Table 55

<table>
<thead>
<tr>
<th>$s_{ij}$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
<th>$g_4$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Now the coefficients of $\sigma_s^2$ in the expected values of the squares and products of sums associated with $a_k$, $g_k$, and $d_j$ are calculated. For example, the coefficient for $a_1 a_1$ is $6^2 + 6^2 + 6^2 + 9^2 + 8^2 + 6^2 = 289$. The coefficient for $a_1 a_2$ is $6(7) + \ldots + 6(6) = 231$ and similarly for all combinations of the parameters. The complete set of coefficients is presented in Table 56.

Table 56

Coefficient of $\sigma_s^2$ in $E(\text{Products of Associated Sums})$

<table>
<thead>
<tr>
<th></th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$g_1$</th>
<th>$g_2$</th>
<th>$g_3$</th>
<th>$g_4$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>289</td>
<td>255</td>
<td>102</td>
<td>169</td>
<td>135</td>
<td>138</td>
<td>114</td>
<td>165</td>
<td>147</td>
<td>118</td>
</tr>
<tr>
<td>$a_2$</td>
<td>231</td>
<td>100</td>
<td>149</td>
<td>117</td>
<td>120</td>
<td>118</td>
<td>138</td>
<td>122</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>$g_1$</td>
<td></td>
<td>99</td>
<td>40</td>
<td>28</td>
<td>35</td>
<td>103</td>
<td>25</td>
<td>49</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>$g_2$</td>
<td></td>
<td></td>
<td>170</td>
<td>63</td>
<td>45</td>
<td>64</td>
<td>148</td>
<td>81</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>$g_3$</td>
<td></td>
<td></td>
<td></td>
<td>129</td>
<td>32</td>
<td>16</td>
<td>49</td>
<td>123</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>$g_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>146</td>
<td>49</td>
<td>81</td>
<td>16</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>$d_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>129</td>
<td>40</td>
<td>28</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>$d_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>155</td>
<td>63</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>$d_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>146</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>$d_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>114</td>
<td></td>
</tr>
</tbody>
</table>

It is understood that there are entries to the left of the diagonal in Table 56 symmetric to those to the right of the diagonal. Then $k_1$ equals the product of each of the entries in this table by the corresponding ones in the invert matrix of Table 53. This product equals $53.51$. Therefore,

$$\widehat{\sigma_s^2} = \frac{1}{78-53.51} \left[ 9.37 - 2(-0.506) \right] = 0.341.$$

3. Estimation of $\sigma_m^2$

$$\widehat{\sigma_m^2} = \frac{1}{k_1} \left[ R(\mu, g, m, s, a) - R(\mu, g, s, a) - k_2 \sigma_e^2 \right].$$

$k_1 = n \ldots$ - coefficient of $\sigma_m^2$ in $R(\mu, g, s, a)$

$k_2 = p - 1 = 3$. 

- 188 -
\( k_1 \) is computed by setting up a table showing the coefficients of \( n_j \) in the expectation of the sums associated with \( S_{ij} \) and \( a_k \). Then the coefficients of \( \sigma^2 \) in the expectation of the squares and cross products among these sums is computed and these coefficients multiplied by corresponding elements of the inverse matrix of the coefficients in the equations of Table 47 from which \( m_j \) have been deleted. The sum of these products is 41.32.

Therefore, \( \delta_2 = \frac{1}{78-41.32} \left[ 6.45 - 3(0.506) \right] = 0.134. \)

4. Estimation of \( \delta_2^2 \)

\[
\delta_2^2 = \left[ R(\mu, g, m, a) - R(\mu, m, a) - k_2 \delta_2^2 - k_3 \delta_2^2 \right]/k_1.
\]

\[
k_1 = 2n \ldots \text{ coefficient of } \delta^2 \text{ in expected value of } R(\mu, m, a).
\]

\[
k_2 = \text{coefficient of } \delta_2^2 \text{ in expected value of } R(\mu, g, m, a)
\]

\[
k_3 = p - 1 = 3.
\]

\[
k_1 = 2(78) - 105.61 = 50.19
\]

\[
k_2 = 53.51 - 27.84 = 25.67.
\]

Therefore, \( \delta_2^2 = \left[ 40.13 - 25.67(0.341) - 3(0.506) \right]/50.19 = 0.595. \)
APPENDIX II. NUMERICAL EXAMPLE OF ANALYSIS OF BALANCED SINGLE CROSS DESIGN

Assume that there are five lines tested and that two experimental units per reciprocal cross are tested. That is, \( p = 5, n = 2 \). The sums of the two experimental units for each of the reciprocal crosses are shown in Table 57.

<table>
<thead>
<tr>
<th>Line of Sire</th>
<th>Line of Dam 1</th>
<th>Line of Dam 2</th>
<th>Line of Dam 3</th>
<th>Line of Dam 4</th>
<th>Line of Dam 5</th>
<th>Sums</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27</td>
<td>28</td>
<td>23</td>
<td>17</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>26</td>
<td>22</td>
<td>15</td>
<td></td>
<td>92</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>21</td>
<td>16</td>
<td>14</td>
<td></td>
<td>79</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>18</td>
<td>15</td>
<td>14</td>
<td></td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>14</td>
<td>16</td>
<td>17</td>
<td></td>
<td>65</td>
</tr>
</tbody>
</table>

\[ \sum y^2_{ijkl} = 4301. \]

A. Least Squares Estimates

Utilizing the formulas given in Section II H 1, the estimates are as follows:

\[ \hat{\mu} = \frac{1}{5(4)(2)} (401) = 10.025. \]

\[ \hat{\xi}_1 = \frac{1}{5(3)(2)} \left[ \frac{4(95) + 98 - 401}{2} \right] = 2.5667 \]

\[ \hat{\xi}_2 = 1.5667 \]

\[ \hat{\xi}_3 = 0 \]

\[ \hat{\xi}_4 = -1.4333 \]

\[ \hat{\xi}_5 = -2.7000 \]
It will be noted that \( \hat{e}_1 + \ldots + \hat{e}_5 = 0 \).

\[
\hat{m}_1 = \frac{1}{5} (2) (98 - 95) = .3000 \quad \hat{m}_4 = .8000
\]
\[
\hat{m}_2 = -1.2000 \quad \hat{m}_5 = -.5000
\]
\[
\hat{m}_3 = .6000
\]

\[
\hat{s}_{12} = \frac{1}{48} \left[ 12 (29 + 27) - 4 (95 + 92 + 98 + 80) + 802 \right] = .2917
\]

\[
\hat{s}_{13} = .9583 \quad \hat{s}_{25} = -.7917
\]
\[
\hat{s}_{14} = -.2083 \quad \hat{s}_{34} = -1.5417
\]
\[
\hat{s}_{15} = -1.0417 \quad \hat{s}_{35} = .1250
\]
\[
\hat{s}_{23} = .4583 \quad \hat{s}_{45} = 1.7083
\]
\[
\hat{s}_{24} = .0417
\]

It will be noted that the sum of the \( \hat{s}_{ij} \) over each line equals zero.

\[
\hat{r}_{12} = \frac{1}{20} \left[ 5 (27 - 29) + 98 - 80 - 95 + 92 \right] = .25
\]

Similarly, the other \( \hat{r}_{ij} \) are:

\[
13 = -.15 \quad 35 = .05
\]
\[
14 = -.25 \quad 41 = .25
\]
\[
15 = .15 \quad 42 = 0
\]
\[
21 = -.25 \quad 43 = -.15
\]
\[
23 = .35 \quad 45 = -.10
\]
\[
24 = 0 \quad 51 = -.15
\]
\[
25 = -.10 \quad 52 = .10
\]
\[
31 = .15 \quad 53 = -.05
\]
\[
32 = -.35 \quad 54 = .10
\]
\[
34 = .15
\]

It will be noted that the sum over each line of sire and over each line of dam equals zero and that \( \hat{r}_{ij} = -\hat{r}_{ji} \).

If it is assumed that \( m_j = 0 \), the estimates of \( g_i \) are as follows:
\[ \hat{e}_1 = \frac{1}{50} \left[ 5 (98 + 95) - 802 \right] = 2.7167. \]

Similarly,
\[ \hat{e}_2 = .9667 \quad \hat{e}_4 = -1.0333 \]
\[ \hat{e}_3 = .3000 \quad \hat{e}_5 = -2.9500 \]

Under the hypothesis the \( m_j = 0 \), \( \hat{r}_{12} = \frac{1}{3} (27 - 29) = -.50 \). Similarly, the other \( \hat{r}_{ij} \) are as follows:

\[
\begin{align*}
13 & = 0 & 35 & = -.50 \\
14 & = 0 & 41 & = 0 \\
15 & = -.25 & 42 & = -1.00 \\
21 & = .50 & 43 & = -.25 \\
23 & = 1.25 & 45 & = -.75 \\
24 & = 1.00 & 51 & = .25 \\
25 & = .25 & 52 & = -.25 \\
31 & = 0 & 53 & = .50 \\
32 & = -1.25 & 54 & = .75 \\
34 & = .25 & \\
\end{align*}
\]

It will be noted that \( \hat{r}_{ij} = -\hat{r}_{ji} \) and that the sum over lines, disregarding the sex, equals zero, but that the sum over line of sire and over line of dam \( \not= 0 \).

B. Tests of Hypotheses

1. Test of the hypothesis that \( r_{ij} = 0 \) assuming \( m_j \not= 0 \)

\[ \frac{1}{\hat{e}_{ij}} \sum_{\hat{r}_{ij}^2} y_{ij}^2 = 4286.50 \]
\[ \frac{1}{\hat{e}_{ij}} \sum_{\hat{r}_{ij}^2} (y_{ij}^* + y_{ij}^*)^2 = 4271.25 \]
\[ \sum_{i=1}^{j} (y_{i..} + y_{.j}) \hat{e}_1 + \sum_{j} y_{.j} \hat{e}_j = 193 (2.5667) + \ldots + 98 (0.3000) + \ldots = 232.0171 \]
\[
\sum_{i=j} \left( y_{i..} + y_{..j} \right) \hat{e}_{ij} = 193 (2.7167) + \ldots = 218.1171.
\]

The sum of squares for \( R = 4286.50 - 4271.25 - 232.0171 \times 218.1171 = 1.35 \), degrees of freedom = \( \frac{1}{2} (4)(3) = 6 \). It will be noted that this sum of squares is equal to \( \sum_{ij} \hat{r}_{ij} y_{ij} \).

The error sum of squares = \( \sum_{ijk} y^2_{ijk} - \frac{1}{n} \sum_{ij} y^2_{ij} = 4301 - 4286.50 = 14.50 \), degrees of freedom = 5 (4)(1) = 20. The analysis of variance is

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among R</td>
<td>6</td>
<td>1.35</td>
<td>.225</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>14.50</td>
<td>.725</td>
</tr>
</tbody>
</table>

2. **Test of the hypothesis that \( a_{ij} = 0 \) assuming \( m_j \neq 0 \)**

\[
\frac{1}{2n} \sum_{ij} (y_{ij} + y_{..j})^2 = 4271.25
\]

\[
\sum_{i=j} \left( y_{i..} + y_{..j} \right) \hat{e}_{ij} = 218.1171
\]

\[
\frac{1}{n} \sum_{i=j} y^2_{ij} = \frac{1}{40} (401)^2 = 4020.0250
\]

The sum of squares for \( S = 4271.25 - 218.1171 - 4020.0250 = 33.1079 \) with 5 degrees of freedom. It will be noted that this sum of squares is the same as \( \sum_{ij} \hat{e}_{ij} (y_{ij} + y_{..j}) \).

The error sum of squares (assuming \( r_{ij} = 0 \)) = 4301 - 4271.25 + 218.1171 - 232.0171 - 4020.0250 = 15.85 with \( \frac{1}{4} (4)(20 - 5 - 2) = 26 \) degrees of freedom. The analysis of variance is

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among S</td>
<td>5</td>
<td>33.11</td>
<td>6.622</td>
</tr>
<tr>
<td>Error</td>
<td>26</td>
<td>15.85</td>
<td>.610</td>
</tr>
</tbody>
</table>

3. **Test of the hypothesis that \( m_j = 0 \)**

\[
\sum_{i=j} \left( y_{i..} + y_{..j} \right) \hat{e}_{ij} + \sum_{j} y_{..j} \hat{m}_j = 232.0171
\]
\[
\bar{y}_{i=j} = 218.1171.
\]

The sum of squares for \( M = 232.0171 = 218.1171 = 13.9000 \) with 4 degrees of freedom. The error sum of squares (assuming \( r_{ij} = 0, s_{ij} \neq 0 \)) = 15.85 with 26 degrees of freedom. The analysis of variance is

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among ( M )</td>
<td>4</td>
<td>13.90</td>
<td>3.475</td>
</tr>
<tr>
<td>Error</td>
<td>26</td>
<td>15.85</td>
<td>0.610</td>
</tr>
</tbody>
</table>

4. Test of the hypothesis that \( g_i = 0 \) assuming \( m_j \neq 0 \)

\[
\bar{y}_{i=j} = 232.0171
\]

\[
\frac{1}{6} \sum_{j} y_{j}^2 = \frac{1}{6} (98^2 + \ldots + 60^2) = 4114.1250
\]

\[
\frac{1}{40} (401)^2 = 4020.0250.
\]

Sum of squares among \( G = 232.0171 - 4114.1250 + 4020.0250 = 137.9171 \) with 4 degrees of freedom. The error sum of squares, assuming \( r_{ij} = 0, s_{ij} \neq 0 \) = 15.85 with 26 degrees of freedom. The analysis of variance is as follows:

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among ( G )</td>
<td>4</td>
<td>137.92</td>
<td>34.480</td>
</tr>
<tr>
<td>Error</td>
<td>26</td>
<td>15.85</td>
<td>0.610</td>
</tr>
</tbody>
</table>

C. Estimation of Variance Components

Utilizing the formulas for the expected values of the sums of squares for \( G, M, S, R, \) and \( E \) given in Section II I 1 and equating the expectations to the computed sums of squares, the equations of Table 58 are obtained for estimating the variances.
Table 58

Equations for Estimating Variance Components

<table>
<thead>
<tr>
<th></th>
<th>$\delta_g^2$</th>
<th>$\delta_m^2$</th>
<th>$\delta_e^2$</th>
<th>$\delta_r^2$</th>
<th>$\delta_s^2$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>30</td>
<td>0</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>137.92</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>13.90</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>33.11</td>
</tr>
<tr>
<td>R</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>6</td>
<td>1.35</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>14.50</td>
</tr>
</tbody>
</table>

The solution is

$$\delta_g^2 = 4.03$$

$$\delta_m^2 = .65$$

$$\delta_s^2 = 1.60$$

$$\delta_r^2 = -.25$$

$$\delta_e^2 = .72$$

D. Estimation of $\delta_{gm}$

$$\mathbb{E} \left[ \xi \hat{g}_{i_1 m_1} \right] = (p-1) \delta_{gm} - \frac{(p-1)}{pn} \delta_e^2$$

$$\mathbb{E} \left[ \xi \hat{g}_{i_1 m_1} \right] = -.9067$$

$$\hat{\delta}_e^2 = .72$$

Therefore, $\hat{\delta}_{gm} = \frac{1}{p-1} \left[ \mathbb{E} \left[ \xi \hat{g}_{i_1 m_1} \right] + \frac{p-1}{pn} \hat{\delta}_e^2 \right] \left[ -.9067 + .2880 \right] = -.15.$

E. Construction of Indexes for Selecting Lines and Crosses

The least squares estimates and the estimates of variances and covariances can now be used to construct various indexes. The expected values needed are as follows:
\[
\hat{E}_g^2 = \frac{p-1}{p} \delta_g^2 + \frac{p-1}{p(p-2)} \delta_s^2 + \frac{(p-1)^2}{p^2(p-2)n} \delta_e^2 = 3.224 + 0.427 + 0.077 = 3.728
\]

\[
\hat{E}_{g_1} = -\frac{1}{p-1} \hat{E}_{g_1}^2 = -0.932
\]

\[
\hat{E}_{g_1}^2 = \frac{p-1}{p} \sigma_{gm} - \frac{p-1}{p^2 n} \delta_e^2 = -0.120 - 0.058 = -0.178
\]

\[
\hat{E}_{g_1} = -\frac{1}{p-1} \hat{E}_{g_1}^2 = 0.044
\]

\[
\hat{E}_{s_1}^2 = 0
\]

\[
\hat{E}_{g_1} = \frac{p-1}{p-2} \delta_g^2 = 3.224
\]

\[
\hat{E}_{g_1} = -\frac{1}{p} \delta_g^2 = -0.806
\]

\[
\hat{E}_{s_1}^2 = \frac{1}{p} \delta_s^2 = 0.320
\]

\[
\hat{E}_{g_1}^2 = \frac{p-1}{p} \sigma_{gm} = -0.120
\]

\[
\hat{E}_{g_1} = -\frac{1}{p} \sigma_{gm} = 0.030
\]

\[
\hat{E}_{m_1}^2 = \frac{p-1}{p} \delta_m^2 + \frac{2(p-1)}{p^2 n} \delta_e^2 = 0.520 + 0.115 = 0.635
\]

\[
\hat{E}_{m_1} = 0
\]

\[
\hat{E}_{s_1}^2 = 0
\]

\[
\hat{E}_{g_1} = \frac{p-1}{p} \delta_g = -0.120
\]

\[
\hat{E}_{m_1}^2 = \frac{p-1}{p} \sigma_m^2 = 0.520
\]

\[
\hat{E}_{s_1}^2 = 0
\]

\[
\hat{E}_{s_1}^2 = \frac{p-3}{p-1} \delta_s^2 + \frac{p-3}{2(p-1)n} \delta_e^2 = 0.800 + 0.090 = 0.890
\]
\[ E_{ij} s_{ij} = \frac{p-3}{p-1} \sigma^2 = .800 \]

\[ E_{ij} s_{ij} = E_{ij} m_j = 0 \]

1. Selection for line of male

The true value, \( T = g_1 \). The appropriate index is

\[ I = b_1 \hat{g}_1 + b_2 \hat{m}_1 \]

The equations which need to be solved to obtain the index are

\[
\begin{array}{ccc}
\hat{b}_1 & \hat{b}_2 & s_i \\
\hline
\hat{E}_{g1} & \hat{E}_{g1m1} & \hat{E}_{g1g1} \\
\hat{E}_{g1m1} & \hat{E}_{m1} & \hat{E}_{m1g1} \\
\end{array}
\]

Substituting the numerical values of the expectations in this set of equations gives

\[
\begin{array}{ccc}
\hat{b}_1 & \hat{b}_2 & s_i \\
\hline
3.728 & -.178 & 3.224 \\
-.178 & .635 & -.120 \\
\end{array}
\]

The solution is \( \hat{b}_1 = .867, \hat{b}_2 = .054 \)

2. Selection for line of female

\[ T = g_j + m_j \]

\[ I = b_1 \hat{g}_j + b_2 \hat{m}_j \]

The equations to be solved are
Substitution of the numerical values of the expectation in the above equations gives the following set.

<table>
<thead>
<tr>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.728</td>
<td>-.178</td>
</tr>
<tr>
<td>-.178</td>
<td>.635</td>
</tr>
<tr>
<td>3.104</td>
<td>.400</td>
</tr>
</tbody>
</table>

The solution is $b_1 = .374$, $b_2 = .375$.

3. Selection for a specific cross

$$T = g_1 + g_j + m_j + s_{ij}$$

$$I = b_1 \hat{g}_1 + b_2 \hat{g}_j + b_3 \hat{m}_j + b_4 \hat{s}_{ij}$$

The equations for computing the $b$'s are as follows:

<table>
<thead>
<tr>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{E}_g^2$</td>
<td>$\hat{E}_g^m_j$</td>
<td>$\hat{E}<em>g^{s</em>{ij}}$</td>
<td>$\hat{E}_g^T$</td>
</tr>
<tr>
<td>$\hat{E}_g^2$</td>
<td>$\hat{E}_g^m_j$</td>
<td>$\hat{E}<em>g^{s</em>{ij}}$</td>
<td>$\hat{E}_g^T$</td>
</tr>
<tr>
<td>$\hat{E}_m^2$</td>
<td>$\hat{E}<em>m^{s</em>{ij}}$</td>
<td>$\hat{E}_m^T$</td>
<td></td>
</tr>
<tr>
<td>$\hat{s}_{ij}$</td>
<td>$\hat{s}_{ij}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The above equations reduce to the following form when the numerical values of the expectations are utilized:
The solution is $b_1 = .979$, $b_2 = .989$, $b_3 = .887$, $b_4 = .899$. 