have ranged from 110 to 168 lb., averaging 145 lb.

With the exception of two lighter weight calves, normal delivery has been possible. Four have been delivered by Caesarean section. Of these none has survived more than a few hours. All other calves were dismembered. As a result, the future reproductive capacity of these cows has been impaired greatly. During the ninth month of pregnancy, there appears to be a complete absence of all physical changes normally occurring.

Pedigree analyses indicate that all calves manifesting this anomaly are homozygous for an autosomal recessive gene. It is concluded tentatively that prolonged gestation is caused by an hormonal imbalance between the fetus and mother, when the fetus is of the mutant genotype. This unique genetic material should prove valuable for certain physiological studies.

P3 Estimation of Changes in Herd Environment. C. R. HENDERSON, Cornell University, Ithaca, N. Y.

Accurate appraisals of the results of breeding programs and most efficient estimates of breeding values of individuals whose own records and whose relatives' records were made in several different years require quantitative measures of the effects of changing herd environment. Least squares or modified least squares methods for obtaining such measures give biased estimates when records of cows culled from the herd are either above or below the herd average. This bias results from the lack of perfect repeatability of records. In contrast, the method of maximum likelihood automatically takes into account incomplete repeatability and annual culling levels and utilizes all of the records in such a way as to obtain the most precise estimates possible of the yearly environmental effects.

The maximum likelihood method has been utilized to obtain annual correction factors for several New York dairy herds. The method is illustrated with data from one of these herds and less accurate but less laborious modifications are described. Examples are given of the use of the correction factors for estimating the genetic improvement in the herd, predicting the breeding values of cows and evaluating sire proofs. Application of the method to computation of age correction factors also is discussed.

P4 The Number of Proved Sons Necessary to Evaluate the Transmitting Ability of a Sire. W. E. WASHBON AND W. J. TYLER, West Virginia University, Morgantown.

One hunderd seventy-four Holstein sires with eight or more D.H.I.A. proved sons were studied to determine the least number of proved sons necessary to estimate most accurately the performance of those to be proved later. The data included average butterfat production of the proved sons' daughters, average difference of daughters' production as compared to their dams and per cent of proved sons that maintained or increased butterfat production in the herds in which they were used. Averages of the first three to ten proved sons, respectively, were compared with averages of the following three, five and ten sons.

Highly significant correlations ($r = 0.35$ to 0.65) indicated that the average butterfat production of the daugthers of the first three proved sons was nearly as accurate as data on more sons in estimating the average butterfat production of the daughters of the next three, five or ten proved sons of a sire.

Similarly, the significant correlations ($r = 0.24$ to 0.40) for the sons' daughters' increase or decrease in butterfat production from their dams indicated that data on the first three or four proved sons were nearly as accurate as data on a larger number in predicting what might be expected from the next three, five or ten proved sons in this respect.

For per cent of sons improving production the correlations were significant when the first four, five and six sons were compared with the next ten sons ($r = 0.30$).

A sire's future granddaughters' butterfat production and its difference from their dams' production apparently can be estimated nearly as well from the performance of the first three or four sons as from a larger number. The per cent of a sire's future proved sons that likely will improve production is more reliable if the prediction is based upon the performance of at least the first four or five proved sons.

P5 Calf Mortality, Sex Ratio and Incidence of Twinning in Two University of Minnesota Herds. K. MILLER AND L. GILMORE, Minn. Agr. Expt. Station, St. Paul.

The University calf records at Grand Rapids and St. Paul have been analyzed with respect to prenatal and postnatal mortality for the first 6 mo. and for other information such as the sex ratio and incidence of multiple births.

At University Farm for the 12 years 1934 to 1945, inclusive, 50 of the 592 births, among Guernseys, Holstein-Fresians and Jerseys, were born dead. Ninty-one died during the first month, 36 more by the end of the sixth month

ESTIMATION OF CHANGES IN HERD ENVIRONMENT

C. R. Henderson
Cornell University

One of the important problems in dairy cattle breeding is the estimation of year to year fluctuations or trends in environmental factors influencing milk production. An examination of herd averages with respect to changes from one year to another tells nothing at all about the causes of these changes. Increased production may be the result of improved feeding and management, of improved genetic merit, or of some combination of changed environment and genotypes. Consequently, if the effectiveness of a breeding program is to be appraised, it is necessary that the observed changes in production be separated into environmental and genetic components. Also most effective selection among individuals who made their records and whose relatives made their records in different years requires that these records be corrected for any year to year differences in environment.

Accounting for environmental changes is particularly important in appraising bull proofs. Since the dams' records are usually made several years earlier than are the daughters' records, any very appreciable change in environment can seriously bias the bull proof. Attempts have been made to get around this difficulty by using only concurrent records in cases where there is reason to suspect that there have been some rather marked changes in feeding and management practices between the time when most of the dams made their records and the time when most of the daughters made their records. The difficulty with this method is that many of the dams leave the herd before their daughters make their records, and consequently a great deal of information must be sacrificed. One way which might be suggested to obviate partially the difficulty due to small numbers is to utilize records of all cows who were in production in both of the years under question. For example, a dam made a record in 1946 and her daughter a record in 1949 and we wish to take into account possible changes in environment between these two years. We could assemble the records of all cows in the herd who made records in both 1946 and 1949 and compute the difference between the average production of this selected group of cows in each of the two years under question.

This method for correcting daughter-dam comparisons seems to offer a logical method for solving the general problem of estimating changes in herd environment. That is, the change from year 1 to year 2 would be the difference between the average of these two years with respect to the selected group of cows who made records in both of these years. Similarly the change from year 2 to year 3 would be the difference in average production for these two years in a second selected group of cows who made records in both years 2 and 3. This procedure could be carried on for any number of years desired. It is obvious that this is not the most efficient possible method of estimating changes, since the change from year 1 to year 3 can be estimated not only by comparison of the year 1 versus year 2 with the year 2 versus year 3 estimates, but also by comparisons within the group of cows with records in both years 1 and 3. It can be seen that there are many possible combinations of comparisons and the problem arises as to how best to weight them.

The method of least squares is a computational procedure which enables one properly to weight the various comparisons. Although it will be shown that this method may yield seriously biased estimates, it is described here because it is frequently used in the solution of problems like this one involving unequal subclass numbers in factorial classifications and is therefore familiar to some workers. Also the maximum likelihood method presented in this paper closely resembles the least squares procedure. A model applicable to the least squares method is as follows:

$y_{ij}=\mu+a_i+c_j+e_{ij}$, where y_{ij} is the record made by the j^{th} cow in the i^{th} year, μ is the population mean, a_i is the environmental effect of the i^{th} year, c_j is the real producing ability of the j^{th} cow, and e_{ij} is a random error associated with the record of the j^{th} cow made in the i^{th} year. The errors are assumed to be uncorrelated, and in most cases they are assumed to be normally distributed. Essentially all this model says is that the record of a cow can be at least approximately expressed as the sum of the population mean, a cow effect, an environmental effect peculiar to a particular year, and an error which is presumably the sum of many different environmental effects impinging upon a particular cow in a particular year. The real producing ability of the cow is a function of her genotype and of permanent environmental factors peculiar to her. The problem is to assign values to μ, a_i, and c_j in such a way as to minimize the residual sum of squares. The solution to the following set of simultaneous equations accomplishes this objective.

$$n_{..}\hat{\mu}+\sum_i n_{i.}\hat{a}_i+\sum_j n_{.j}\hat{c}_j=\sum_{ij}\sum y_{ij}$$

$$n_{i.}\hat{\mu}+n_{i.}\hat{a}_i+\sum_j n_{ij}\hat{c}_j=\sum_j y_{ij} \quad \text{(One such equation for each of the } a_i\text{)}$$

$$n_{.j}\hat{\mu}+\sum_i n_{ij}\hat{a}_i+n_{.j}\hat{c}_j=\sum_i y_{ij} \quad \text{(One such equation for each } c_j\text{)},$$ where $n_{ij}=1$ if the j^{th} cow made a record in the i^{th} year and $=0$ if the j^{th} cow did not make a record in the i^{th} year, $n_{..}$ = the total number of records, $n_{i.}$ = the number of records made in the i^{th} year, and $n_{.j}$= the number of records made by the j^{th} cow. The solution to these equations is not as formidable as first appears for $\hat{\mu}+\hat{c}_j=\frac{1}{n_{.j}}(\sum_i y_{ij}-\sum_i n_{ij}\hat{a}_i)$. Utilizing this fact, the equations can be reduced to ones involving only the $\hat{a}_i$. Then after imposing the restriction $\sum_i \hat{a}_i=0$, the $\hat{a}_i$ can be solved.

As was stated above, the least squares estimate, and the same can be said of comparisons of yearly averages of cows selected because they had records in two specified years, can give seriously biased results. The bias is of this sort. If the cows which are culled from the herd have average records below the herd mean, the environment will appear to be getting worse and worse if in fact there has been no change in environment. Similarly if the environment has been improving, improvement in environment appears less than it really has been. If the environment has been getting worse, the deterioration in environment will appear worse than it actually has been. Conversely, if the cows culled from the herd are of above average production, the environment will seem to have improved more or to have deteriorated less than is actually the case.

The reason for the bias in the least squares method is the incomplete repeatability of dairy records. It is a well known fact that a group of cows selected because their records were above the herd average during the past one or more years are likely to produce less far above the herd average next year. Conversely cows selected for below average production are likely to be less far below the herd average in their succeeding lactations. The reason for this is that cows with above average production, more frequently than not, receive an above average temporary environmental contribution. Since this temporary environmental factor does not carry over to succeeding lactations, these later records are likely to fall below the earlier ones. The crucial point is that least squares and similar methods are essentially year to year comparisons of records of the survivors of each year's culling. Therefore, if the survivors of culling made above herd average records prior to the culling of certain of their mates, we should expect these survivors' records to be less in succeeding years and consequently to make it appear that the environment is becoming poorer from year to year.

What is needed then is some method by which knowledge of repeatability of dairy records and of the year to year culling levels can be utilized to adjust the estimates of the yearly environmental effects. The method of maximum likelihood now to be described does all of this automatically in a set of simultaneous equations which are little more complex than the least squares equations. In addition the method effects faster genetic progress through selection on own performance than does any other method for utilizing annual production records.

Now let us modify the model used for the least squares analysis as follows:

$y_{ijk}=\mu+a_i+b_j+c_{jk}+e_{ijk}$, where y_{ijk} is the record made in the i^{th} year by the k^{th} cow of the j^{th} group of cows, μ is the population average, a_i is the environmental effect of the i^{th} year, b_j is the average real producing ability of the j^{th} group of cows, c_{jk} is the real producing ability of the k^{th} cow of the j^{th} group, and e_{ijk} is a random environmental effect peculiar to the individual record. The new element in this model as compared to the least squares model is the b_j. It represents the average real producing ability of a particular group of cows. It might, for example, represent the daughters of a particular bull or a set of cows born within a specified period. Now let us make the additional assumptions that the c_{jk} are normally and independently distributed with mean 0 and variance σ_c^2, that the e_{ijk} are normally and independently distributed with mean 0 and variance σ_e^2, and that the c's and e's are uncorrelated. These assumptions mean that the cows of a particular group are randomly drawn from a normal population whose mean is $\mu+b_j$ and whose variance is σ_c^2. Furthermore temporary environment is not correlated with real producing ability.

With the assumptions specified above we can now write the joint distribution of the y_{ijk} and the c_{jk} and can proceed to compute values of μ, a_i, b_j, and c_{jk} which will maximize the probability of obtaining the sample of records actually at hand. The distribution is

$$L=\prod_{ijk}\frac{1}{\sqrt{2\pi\sigma_e^2}}e^{-\frac{1}{2\sigma_e^2}(y_{ijk}-\mu-a_i-b_j-c_{jk})^2}\prod_{jk}\frac{1}{\sqrt{2\pi\sigma_c^2}}e^{-\frac{c_{jk}^2}{2\sigma_c^2}}$$

The maximizing values are the solution to the following set of simultaneous equations:

$$n_{...}\tilde{\mu}+\sum_i n_{i..}\tilde{a}_i+\sum_j n_{.j.}\tilde{b}_j+\sum\sum_{jk} n_{.jk}\tilde{c}_{jk}=\sum\sum\sum_{ijk} y_{ijk}$$

$n_{i..}(\tilde{\mu}+\tilde{a}_i)+\sum_j n_{ij.}\tilde{b}_j+\sum\sum_{jk} n_{ijk}\tilde{c}_{jk}=\sum\sum_{jk} y_{ijk}$, and similarly for all a's.

$n_{.j.}(\tilde{\mu}+\tilde{b}_j)+\sum_i n_{ij.}\tilde{a}_i+\sum_k n_{.jk}\tilde{c}_{jk}=\sum\sum_{ik} y_{ijk}$, and similarly for all b's.

$n_{.jk}(\tilde{\mu}+\tilde{b}_j)+\sum_i n_{ijk}\tilde{a}_i+(n_{.jk}+\frac{\sigma_e^2}{\sigma_c^2})\tilde{c}_{jk}=\sum_i y_{ijk}$, and similarly for all c's.

$n_{ijk}=1$ if the jk^{th} cow has a record in the i^{th} year and $=0$ otherwise. A dot in the subscript denotes summation over that particular subscript. That is, $n_{...}$ refers to the total number of records, $n_{i..}$ to the number of records in the i^{th} year, $n_{.j.}$ to the number of records made by the cows of the j^{th} group, $n_{.jk}$ to the number of records made by the k^{th} cow of the j^{th} group, and $n_{ij.}$ to the number of records made by the j^{th} group in the i^{th} year.

It will be noted that $\frac{\sigma_e^2}{\sigma_c^2}$ appears in certain coefficients. This ratio is closely related to repeatability of dairy records as defined by Lush, and concerning the size of which several studies have been made. If we denote repeatability by $r=\frac{\sigma_c^2}{\sigma_c^2+\sigma_e^2}$, then $\frac{\sigma_e^2}{\sigma_c^2}$ can be written as $\frac{1-r}{r}$ For example, if we take $r=.40$, $\frac{\sigma_e^2}{\sigma_c^2}=1.5$.

Solving for $\tilde{c}_{jk}$ in the above equations we get

$$\tilde{c}_{jk}=\frac{\sigma_c^2}{n_{.jk}\sigma_c^2+\sigma_e^2}(\sum_i y_{ijk}-n_{.jk}\tilde{\mu}-n_{.jk}\tilde{b}_j-\sum_i n_{ijk}\tilde{a}_i).$$

Consequently we can substitute these expressions for $\tilde{c}_{jk}$ in the μ, a, and b equations and thereby eliminate the $\tilde{c}_{jk}$. Then we have a set of equations in which $\tilde{\mu}+\tilde{b}_j$ can be expressed in terms of the $\tilde{a}_i$ and certain observed records in which by proper substitution the equations can be reduced to one involving only the $\tilde{a}_i$.

Let us look at the expression for $\tilde{c}_{jk}$ and now assume that μ is known and that all a_i and $b_j = 0$. This would be the case if we assume that there have been no changes in the herd environment nor in the genetic merit of the herd and that the herd average is known without error. Then,

$$\tilde{c}_{jk} = \frac{\sigma_c^2}{n_{\cdot jk}\sigma_c^2 + \sigma_e^2} \left(\sum_i y_{ijk} - n_{\cdot jk}\mu\right)$$

$$= \frac{n_{\cdot jk}\, r}{1 + (n_{\cdot jk} - 1) r} \left(\bar{y}_{\cdot jk} - \mu\right)$$

This last expression is exactly equivalent to the method derived by Lush for estimating the real producing ability of a cow. That is, his method is in fact the maximum likelihood method under the assumptions listed above. The method described here utilizes these same principles but permits the assumption of changing herd environment and genotype.

Several different modifications of the model described above are possible. For example, linear or curvilinear regressions could be fit to freshening date and to date of birth. In any case the general procedures are the same. Also if there is some question regarding the applicability of standard age correction factors to a particular herd, the age of the cow at time of freshening can be worked into the model.

Summary and Conclusions

A maximum likelihood solution to the problem of estimating changes in herd environment has been presented. The method utilizes in the most efficient possible way all of the records in a herd, gives estimates which are free from the bias inherent in least squares or similar methods, and automatically yields estimates of the real producing abilities of each of the cows.

The computational procedure is not particularly difficult for those familiar with least squares analyses. A number of computational short-cuts have been developed and have been described in a numerical example which can be obtained by writing the author at Cornell University.

NUMERICAL EXAMPLE OF ESTIMATION OF CHANGES IN HERD ENVIRONMENT

C. R. Henderson
Cornell University

It should be emphasized that the amount of data in the following example is too small to give accurate estimates of changes in herd environment. The numbers of records have been kept low in order to simplify the presentation.

Table 1 shows the mature equivalent records of 35 cows for the years 1946, 1947 and 1948. The cows have been divided into four groups depending upon their date of birth. These groups include a group born before 1944, a group born in 1944, a group born in 1945, and a group born in 1946.

The equations for estimating the changes in environment by means of the method of maximum likelihood can be simplified somewhat in comparison to those given in the paper presented at the 1949 meeting of the American Dairy Science Association. This simplification involves combining μ and b_j into a single parameter. Then the equations are as follows:

$n_{i..}a_i+\sum_j n_{ij.}(\mu+b_j)+\sum\sum_{jk} n_{ijk}c_{jk}=\sum\sum_{jk} y_{ijk}$, and similarly for the other a_i equations.

$\sum_i n_{ij.}a_i+n_{.j.}(\mu+b_j)+\sum_k n_{.jk}c_{jk}=\sum\sum_{ik} y_{ijk}$, and similarly for the other $\mu+b_j$ equations.

$\sum_i n_{ijk}a_i+n_{.jk}(\mu+b_j)+(n_{.jk}+\sigma_e^2/\sigma_c^2)c_{jk}=\sum_i y_{ijk}$, and similarly for the other c_{jk} equations.

a_1, a_2, and a_3 = the environmental effects peculiar to cows freshening in 1946, 1947, and 1948 respectively.

$\mu+b_1$, $\mu+b_2$, $\mu+b_3$, and $\mu+b_4$ = the permanent producing abilities of the sub-populations born before 1944, in 1944, in 1945, and in 1946 respectively.

TABLE 1

MATURE EQUIVALENT RECORDS

Cow	Year of Birth	Year Cow Freshened 1946	1947	1948	Cow Sum
1	Before 1944	343	492	311	1146
2		437	544	482	1463
3		390	355		745
4		264			264
5		410			410
6		384			384
7		405	230		635
8		324	395	347	1066
9		447	369	393	1209
10		381	425	392	1198
11		367	397	317	1081
12		282			282
13		395	370	246	1011
14		344	477		821
15		381	339	342	1062
16		333	299	380	1012
17		263			263
18		376	393	349	1118
19		412	322	291	1025
20		381	272		653
21	1944	389	320	326	1035
22		441	378	274	1093
23		384	410	353	1147
24		514	388	453	1355
25		421	332	332	1085
26	1945		273		273
27			385	265	650
28			343	355	698
29			233	245	478
30			346	323	669
31	1946			254	254
32				233	233
33				304	304
34				427	427
35				372	372
Freshening Year Sum		9468	9087	8366	26921

The cows actually born in any specified year are considered to be a random sample of cows from some infinite sub-population of cows which might have been born in that year in the herd being studied. The sub-population is assumed to be normally distributed with mean $\mu+b_j$ and variance σ_c^2.

c_{11} = the permanent producing ability of the first cow born before 1944, her permanent producing ability being expressed as a deviation from the mean of her sub-population (cows born before 1944). In what is to follow c_{11} will be changed to c_1 to correspond to the cow numbers presented in Table 1. Similarly c_{12} will be changed to c_2, etc.

The individual records are assumed to be subject to a random error, e_{ijk}, which is normally distributed with mean 0 and variance σ_e^2.

In this example, the equation pertaining to a_1 is

$$25a_1+20(\mu+b_1)+5(\mu+b_2)+c_1+c_2+\ldots+c_{25}= 9468.$$

25, the coefficient of a_1, arises from the fact that there were 25 records made by cows freshening in 1946. 20, the coefficient of $\mu+b_1$, comes from the fact that of records made by cows freshening in 1946 there were 20 records made by cows who were born before 1944. 5, the coefficient of $\mu+b_2$, arises from the fact that of records made by cows freshening in 1946, 5 records were made by cows born in 1944. 1, the coefficient of c_1, c_2, etc., comes from the fact that of records made by cows freshening in 1946, the cows involved were 1, 2, ..., 25, each with one record. The right member of the equation, 9468, is the sum of all records started in 1946. The equations pertaining to a_2 and a_3 are obtained in a similar manner and are presented in Table 2.

The equation pertaining to $\mu+b_1$ is

$$20a_1+15a_2+11a_3+46(\mu+b_1)+3c_1+3c_2+\ldots+2c_{20}= 16,848$$

20, the coefficient of a_1, is the number of records made by cows born before 1944 and in which the freshening date was 1946. 15, the coefficient of a_2, is the number of records made by this same group of cows in which the freshening date was 1947. 11, the coefficient of a_3, is the number of records made by the same group of cows in which the freshening date was 1948. 46, the coefficient of $\mu+b_1$, is the total number of records made by cows born before 1944. 3, the coefficient of c_1, is the total number of records made by the first cow born before 1944, and similarly for the coefficients of the other c_{1k}. 16,848, the right member of the equation, is the sum of all records made by cows born before 1944. The equations pertaining to $\mu+b_2$, $\mu+b_3$, and $\mu+b_4$ are set up in a similar manner and are presented in Table 2.

The equation pertaining to c_1 is

$$a_1+a_2+a_3+3(\mu+b_1)+4.5c_1 = 1146$$

1, the coefficient of a_1, a_2, and a_3, results from the fact that cow 1 made records in which the freshening dates were 1946, 1947, and 1948. 3, the coefficient of $\mu+b_1$, results from the fact that cow 1 made 3 records and is a member of the sub-population of cows born before 1944. 4.5, the coefficient of c_1, is the sum of 3, the total number of records made by cow 1, and 1.5, the value of σ_e^2/σ_c^2 used in this example. σ_e^2/σ_c^2 is equal to $(1-r)/r$, where r is a repeatability of butterfat records. Equations for the other c_{jk} are set up in a similar manner.

Table 2 presents the full set of maximum likelihood equations in tabular form. The equations pertaining to a_i and $\mu+b_j$ are read vertically. The column headings denote the equations and the row headings denote the unknowns (parameter estimates). The entries in the table are the appropriate coefficients of the unknowns. The last row contains the right members of the equations. For example, the equation pertaining to $\mu+b_3$ is

$$5a_2+4a_3+9(\mu+b_3)+c_{26}+2c_{27}+2c_{28}+2c_{29}+2c_{30}=2768.$$

The equations pertaining to the c_{jk} are read horizontally, the column headings denoting the unknowns and the right members. The row headings denote the equations. In the next to the extreme right column is a heading, c_{jk}. This means that the only c_{jk} unknown in the equation is the one pertaining to the jk^{th} cow. For example, the equation for cow 30 is

$$a_2+a_3+2(\mu+b_3)+3.5c_{30}= 669.$$

TABLE 2

MAXIMUM LIKELIHOOD EQUATIONS

	a_1	a_2	a_3	$\mu+b_1$	$\mu+b_2$	$\mu+b_3$	$\mu+b_4$	c_{jk}	Right Member
a_1	25	0	0	20	5	0	0	-	-
a_2	0	25	0	15	5	5	0	-	-
a_3	0	0	25	11	5	4	5	-	-
$\mu+b_1$	20	15	11	46	0	0	0	-	-
$\mu+b_2$	5	5	5	0	15	0	0	-	-
$\mu+b_3$	0	5	4	0	0	9	0	-	-
$\mu+b_4$	0	0	5	0	0	0	5	-	-
c_1	1	1	1	3	0	0	0	4.5	1146
c_2	1	1	1	3	0	0	0	4.5	1463
c_3	1	1	0	2	0	0	0	3.5	745
c_4	1	0	0	1	0	0	0	2.5	264
c_5	1	0	0	1	0	0	0	2.5	410
c_6	1	0	0	1	0	0	0	2.5	384
c_7	1	1	0	2	0	0	0	3.5	635
c_8	1	1	1	3	0	0	0	4.5	1066
c_9	1	1	1	3	0	0	0	4.5	1209
c_{10}	1	1	1	3	0	0	0	4.5	1198
c_{11}	1	1	1	3	0	0	0	4.5	1081
c_{12}	1	0	0	1	0	0	0	2.5	282
c_{13}	1	1	1	3	0	0	0	4.5	1011
c_{14}	1	1	0	2	0	0	0	3.5	821
c_{15}	1	1	1	3	0	0	0	4.5	1062
c_{16}	1	1	1	3	0	0	0	4.5	1012
c_{17}	1	0	0	1	0	0	0	2.5	263
c_{18}	1	1	1	3	0	0	0	4.5	1118
c_{19}	1	1	1	3	0	0	0	4.5	1025
c_{20}	1	1	0	2	0	0	0	3.5	653
c_{21}	1	1	1	0	3	0	0	4.5	1035
c_{22}	1	1	1	0	3	0	0	4.5	1093
c_{23}	1	1	1	0	3	0	0	4.5	1147
c_{24}	1	1	1	0	3	0	0	4.5	1355
c_{25}	1	1	1	0	3	0	0	4.5	1085
c_{26}	0	1	0	0	0	1	0	2.5	273
c_{27}	0	1	1	0	0	2	0	3.5	650
c_{28}	0	1	1	0	0	2	0	3.5	698
c_{29}	0	1	1	0	0	2	0	3.5	478
c_{30}	0	1	1	0	0	2	0	3.5	669
c_{31}	0	0	1	0	0	0	1	2.5	254
c_{32}	0	0	1	0	0	0	1	2.5	233
c_{33}	0	0	1	0	0	0	1	2.5	304
c_{34}	0	0	1	0	0	0	1	2.5	427
c_{35}	0	0	1	0	0	0	1	2.5	372
Right Member	9468	9087	8366	16,848	5715	2768	1590	-	-

The solution to the equations in Table 2 is a formidable task as they stand for there are 42 unknowns. It is, however, quite easy to reduce these equations to 7 in number by making use of the fact that c_{jk} can be expressed in terms of the a_i, $\mu+b_j$, and the right member of the c_{jk} equation. For example,

$$c_{30} = \frac{1}{3.5}\left[669-a_2-a_3-2(\mu+b_3)\right] .$$

The computational procedure for making such substitutions is facilitated by preparing a table like that of Table 3. The entries in this table <u>not in parentheses</u> are the same as in the lower portion of Table 2. The entries <u>in parentheses</u> are obtained by dividing each of the entries <u>not in parentheses</u> in a particular row by the entry in that row under the column headed c_{jk}. For example, in the first row of the table, the entry in parentheses under a_1 is 1/4.5 = .2222. The entries under a_2 and a_3 in this row are also 1/4.5. The entry under $\mu+b_1$ is 3/4.5 = .6667. The entry under "Right Member" is 1146/4.5 = 254.67.

Table 4 can now be prepared. It is a new set of equations in which the c_{jk} have been expressed in terms of the other parameters. In these equations the column headings denote the unknowns while the row headings denote the equations. The right column entries denote the right members of the equations. This table is prepared by using information in Tables 2 and 3. For example, the entry in the upper left hand corner of Table 4 is equal to 25-1(.2222)-1(.2222)-1(.2857)-...-1(.2222) = 18.30. That is, this entry is equal to the original coefficient of a_1 in the a_1 equation minus the cross products of the entries in the a_1 column of Table 3. The entry in the a_2 column of the a_1 row of Table 4 is equal to 0-1(.2222)-1(.2222)-1(.2857)-1(.2857)-...-1(.2222)-1(.2222) = -4.70. This entry is simply the coefficient of a_2 in the a_1 equation of Table 2 minus the cross products between the entries <u>not in parentheses</u> in column a_1 and entries <u>in parentheses</u> in column a_2 of Table 3. The entry in row a_1, column a_3 of Table 4 is equal to 0-1(.2222)-1(.2222)-...-1(.2222)-1(.2222) = -3.56. The entry in the $\mu+b_1$ column

TABLE 3

Cow	a_1	a_2	a_3	$\mu+b_1$	$\mu+b_2$	$\mu+b_3$	$\mu+b_4$	c_{jk}	Right Member	
1	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1146	(254.67)
2	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1463	(325.11)
3	1(.2857)	1(.2857)		2(.5714)				3.5	745	(212.86)
4	1(.4000)			1(.4000)				2.5	264	(105.60)
5	1(.4000)			1(.4000)				2.5	410	(164.00)
6	1(.4000)			1(.4000)				2.5	384	(153.60)
7	1(.2857)	1(.2857)		2(.5714)				3.5	635	(181.43)
8	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1066	(236.89)
9	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1209	(268.67)
10	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1198	(266.22)
11	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1081	(240.22)
12	1(.4000)			1(.4000)				2.5	282	(112.80)
13	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1011	(224.67)
14	1(.2857)	1(.2857)		2(.5714)				3.5	821	(234.57)
15	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1062	(236.00)
16	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1012	(224.89)
17	1(.4000)			1(.4000)				2.5	263	(105.20)
18	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1118	(248.44)
19	1(.2222)	1(.2222)	1(.2222)	3(.6667)				4.5	1025	(227.78)
20	1(.2857)	1(.2857)		2(.5714)				3.5	653	(186.57)
21	1(.2222)	1(.2222)	1(.2222)		3(.6667)			4.5	1035	(230.00)
22	1(.2222)	1(.2222)	1(.2222)		3(.6667)			4.5	1093	(242.89)
23	1(.2222)	1(.2222)	1(.2222)		3(.6667)			4.5	1147	(254.89)
24	1(.2222)	1(.2222)	1(.2222)		3(.6667)			4.5	1355	(301.11)
25	1(.2222)	1(.2222)	1(.2222)		3(.6667)			4.5	1085	(241.11)
26		1(.4000)				1(.4000)		2.5	273	(109.20)
27		1(.2857)	1(.2857)			2(.5714)		3.5	650	(185.71)
28		1(.2857)	1(.2857)			2(.5714)		3.5	698	(199.43)
29		1(.2857)	1(.2857)			2(.5714)		3.5	478	(136.57)
30		1(.2857)	1(.2857)			2(.5714)		3.5	669	(191.14)
31			1(.4000)				1(.4000)	2.5	254	(101.60)
32			1(.4000)				1(.4000)	2.5	233	(93.20)
33			1(.4000)				1(.4000)	2.5	304	(121.60)
34			1(.4000)				1(.4000)	2.5	427	(170.80)
35			1(.4000)				1(.4000)	2.5	372	(148.80)

TABLE 4

	a_1	a_2	a_3	$\mu+b_1$	$\mu+b_2$	$\mu+b_3$	$\mu+b_4$	Right Member
a_1	18.30	-4.70	-3.56	8.38	1.67	0	0	3987.81
a_2	-4.70	18.76	-4.70	5.38	1.67	2.31	0	3425.96
a_3	-3.56	-4.70	18.30	3.67	1.67	1.71	3.00	2993.59
$\mu+b_1$	8.38	5.38	3.67	17.43	0	0	0	6315.26
$\mu+b_2$	1.67	1.67	1.67	0	5.00	0	0	1905.00
$\mu+b_3$	0	2.31	1.71	0	0	4.03	0	1233.10
$\mu+b_4$	0	0	3.00	0	0	0	3.00	954.00

of the a_1 row of Table 4 is 20-1(.6667)-1(.6667)-1(.5714)-...-1(.6667)-1 (.5714)=8.38. In a similar manner the entries in the $\mu+b_2$, $\mu+b_3$, and $\mu+b_4$ columns of the b_1 row are obtained. The right member of Table 4 in the a_1 row is 9468-1(254.67)-1(325.11)-...-1(301.11)-1(241.11)=3987.81. In a similar manner each of the other entries of Table 4 can be computed. Since the entries are symmetric about the upper left hand to lower right hand diagonal, it is necessary to compute only the diagonal elements and those to the right of it. The entries to the left of the diagonal can be copied from those already computed for those to the right.

There is a useful check on the accuracy of the entries in Table 4. In each row the sum of the entries under the a_i is equal to the sum of the entries under the $\mu+b_j$. For example in the first row 18.30-4.70-3.56=8.38 +1.67 except for rounding error. A similar check exists for the entries in the right hand column of Table 4. The sum of the entries in the a_i rows is equal to the sum of the entries in the $\mu+b_j$ rows except for rounding errors. That is, 3987.81+3425.96+2993.59=6315.26+1905.00+1233.10+954.00.

Although the equations have now been reduced to 7 in number it is possible to reduce them still further by making use of the fact that $\mu+b_j$ can be expressed in terms of the a_j and the right members. Making use of this fact the equations can be reduced to 3 in number. Table 5 facilitates these substitutions.

The entries in Table 5 correspond to the last four rows of Table 4. The entries <u>in parentheses</u> in Table 5 are the entries <u>not in parentheses</u> divided by the entry under $\mu+b_j$ for that particular row. For example the entry in parentheses in the first column and first row is 8.38/17.43=.4808.

Now Table 6 is set up. It is a set of equations in which both the c_{jk} and $\mu+b_j$ have been expressed in terms of the a_i and the right members. The procedure for doing this is just like that used in preparing Table 4. For example the entry in the a_1 row and a_1 column of Table 6 is 18.30-8.38 (.4808)-1.67(.3340) = 13.71. The entry in the a_2 column of the a_1 row of

Table 6 is -4.70-8.38(.3087)-1.67(.3340)= -7.84. The entry in the right member of the first row is 3987.81-.4808(6315.26)-.3340(1905.00) = 315.16. A useful check on the accuracy of the computations of Table 6 is the fact that the entries in each row sum to 0 over the first 3 columns except for rounding error. For example 13.71-7.84-5.88 = -.01. The right members also add to 0 except for rounding errors. That is, 315.16+133.36-449.88 = -1.36.

TABLE 5

$\mu+b_j$	a_1		a_2		a_3		$\mu+b_j$	Right Member
1	8.38	(.4808)	5.38	(.3087)	3.67	(.2106)	17.43	6315.26
2	1.67	(.3340)	1.67	(.3340)	1.67	(.3340)	5.00	1905.00
3	0	(0)	2.31	(.5732)	1.71	(.4243)	4.03	1233.10
4	0	(0)	0	(0)	3.00	(1.0000)	3.00	954.00

TABLE 6

	a_1	a_2	a_3	Right Member
a_1	13.71	-7.84	-5.88	315.16
a_2	-7.84	15.22	-7.37	133.36
a_3	-5.88	7.37	13.24	-449.88

There is no solution to the equations of Table 6 as they stand inasmuch as the sum of any two of the equations is equal to the third equation with signs changed. This difficulty can easily be got around by imposing the restriction that the sum of the a_i=0, that is $a_1+a_2+a_3 = 0$. This is a perfectly logical restriction inasmuch as the yearly environmental effects are thereby expressed as deviations about the mean of such effects. Making use of this restriction, $a_3= -a_1-a_2$. Consequently the equations of Table 6 can be reduced to those of Table 7. For example, the coefficient of a_1 in the a_1 equation of Table 7 is equal to 13.71-(-5.88)=19.59. The coefficient of a_2 in the a_1 equation is -7.84-(-5.88)= -1.96. The a_3 equation of Table 6 can be deleted since it is redundant.

TABLE 7

	a_1	a_2	Right Member
a_1	19.59	-1.96	315.16
a_2	-.47	22.59	133.36

The solution to the equations of Table 7 is $\tilde{a}_1=17$ and $\tilde{a}_2=6$. Consequently $\tilde{a}_3=-17-6 = -23$. What these figures mean is, that on the basis of the available data, cows freshening in 1946 receive an average environment which results in butterfat production of 17 pounds above the average of the three years studied. Consequently, records made in that year should be corrected by subtracting 17 pounds from each of them. In a like manner records made in 1947 should be corrected by subtracting 6 pounds, and records of cows freshening in 1948 should be corrected by adding 23 pounds.

We can now proceed to obtain the maximum likelihood estimates for the $\mu+b_j$ and the c_{jk}. The former are obtained by substituting the values just obtained for the $\tilde{a}_i$ in the last four equations of Table 4 and solving for the $\mu+b_j$. For example,

$$\widetilde{\mu+b}_1=\frac{1}{17.43}\left[6315.26-17(8.38)-6(5.38)+23(3.67)\right]=357.$$

In a like manner the value of the other $\widetilde{\mu+b}_j$ are 381, 312, and 341. Finally if we substitute in the c_{jk} equations of Table 2 the computed values for the $\tilde{a}_i$ and $\widetilde{\mu+b}_j$, we can obtain the estimates of c_{jk}. For example,

$$\tilde{c}_1=\frac{1}{4.5}\left[1146-1(17)-1(6)-1(-23)-3(357)\right]=17$$

Then the estimate of the real producing ability of the first cow is

$$\widetilde{\mu+b}_1+\tilde{c}_1=357+17=374,$$ and similarly for the other cows.