Genetic influences on the composition of cows' milk

Carl Max Von Krosigk
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/rtd
Part of the Agriculture Commons, and the Animal Sciences Commons

Recommended Citation
Von Krosigk, Carl Max, "Genetic influences on the composition of cows' milk " (1959). Retrospective Theses and Dissertations. 2188.
http://lib.dr.iastate.edu/rtd/2188

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.
GENETIC INFLUENCES ON THE COMPOSITION
OF COWS' MILK

by

Carl Max von Krosigk

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

Major Subject: Animal Breeding

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
of Science and Technology
Ames, Iowa

1959
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>REVIEW OF LITERATURE</td>
<td>3</td>
</tr>
<tr>
<td>NATURE AND SOURCE OF DATA</td>
<td>21</td>
</tr>
<tr>
<td>METHODS OF ANALYSIS AND RESULTS</td>
<td>28</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>120c</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>130</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>135</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>141</td>
</tr>
</tbody>
</table>
INTRODUCTION

A long-overdue increased awareness of their importance recently has focused attention on the non-fat solids content of milk. In Great Britain this attention has received motivation by an increasing percentage of the milk supply falling below their minimum legal requirements for solids-not-fat content. The Continental Europeans apparently have concluded that the protein content deserves more attention. Indeed, in parts of the Netherlands they already have commenced testing cows for and paying producers on the basis of protein content. Several experiment stations in the United States have inaugurated or are contemplating starting experiments to obtain more information on causes of variation and covariation in milk constituents. Of course, dairy chemists and workers concerned with milk secretion have been acquisitive ab initio for knowledge on all phases of milk composition.

Animal fats generally have decreased in volume in the diet during the last few years. This decrease has resulted primarily from increased competition of vegetable fats and the diminishing need for high energy diets in our modern mechanized society. Swine producers already have made progress in producing hog carcasses with higher percentages of lean cuts. Supposedly some effort may be made in the future to alter the composition of milk. At the least, more economic
and nutritional emphasis will be attached to the protein and solids-not-fat contents.

Relevant studies on the role of heredity as a cause of variation in the non-fat solids have been few. If, in fact, protein and solids-not-fat contents and yields partially or wholly supplant butterfat as the primary ingredient of the worth of a dairy cow, the basic statistics necessary for formulating optimum breeding plans and selection schemes will be needed.

The initial purpose of this investigation was to establish the normal composition of cow's milk in Oregon. These data were not collected primarily to estimate genetic parameters. However, considering the present state of knowledge and the current interest in these parameters, it was deemed important to attempt to estimate them anyway. This attempt was made by substituting statistical adjustments to correct for effects which ordinarily would be "balanced out" in a well designed genetic experiment. In one sense this situation was fortunate because it forced a more searching analysis than otherwise would have been necessary.
REVIEW OF LITERATURE

Recently, several literature reviews have been published on the aspects of milk composition considered in this study. Literature on the environmental and physiological factors affecting milk composition has been reviewed by Bailey (1952a), Herrmann (1954), Meiser (1956), Ling (1958), Larson (1958) and others. An earlier review well worth consulting is that by McDowall (1936). Armstrong (1959) summarized information collected in North America on breed differences. Tyler (1958) presented a critical review of the literature on the present knowledge of genetic influences on milk composition.

Because so many reviews are readily available only the reports from the more extensive, or more recent sets of data will be included in this discourse. Neither will fat content be considered, since its variations are well described in such textbooks as Espe and Smith (1952).

Influence of Age

It has become increasingly apparent that in spite of statements to the contrary, such as those by Bonnier and Hansson (1946), Johansson and Claesson (1957) or Johnson (1957), the effect of age can be considerable especially with lactation averages of solids-not-fat or lactose contents.
Using lactation records of 335 Holstein cows, Gowen (1919) found a significant correlation of -.16 between solids-not-fat content and age. Later (1924) he published the corresponding regression (-.024 per cent) of solids-not-fat on year of age. In Gowen's results the regression was not significantly different from linear.

Tocher (1925) derived regressions of solids-not-fat and total nitrogen on age of cow. For 676 samples from the same number of cows the regression for solids-not-fat was -.032 per year. When the total nitrogen is converted to protein content, the average change in protein was -.0076 per cent for each year of age. Albumin nitrogen increased, but casein nitrogen and lactose content decreased as the cows aged.

Bartlett (1934), using paired lactation records by the same cows, observed an apparently linear decrease of .52 in solids-not-fat content from the first to the ninth lactations. Bailey (1952b) reported a substantially linear decrease of .3 in solids-not-fat content from the second to seventh lactations. Bailey used 423 lactation records from the same herd of Dairy Shorthorns on which Bartlett reported.

Bonnier and Hansson (1946) calculated the differences between first and second lactations for protein and lactose at fixed fat percentages. These differences were not signif-
leant, and from this very limited evidence they decided to disregard age differences in their analysis.

Provan (1955) quoted results, obtained by the Milk Marketing Board, from more than 1,700 individual monthly samples from nine farms in England. The solids-not-fat content decreased from one lactation to the next in all herds. The rate of decrease was about .1 for each lactation up to the fourth; then lessened in later lactations.

From lactation records on 814 Ayrshire cows, Waite et al. (1956) observed differences in the constituent contents from the first to the average of the ninth and later lactations of -.21 casein, -.08 crude protein, -.25 lactose and -.34 solids-not-fat.

Robertson et al. (1956) state that adjusting the lactation records for a herd with average age distribution of cows will decrease the total variance within herds by the following percentages; 11 for casein, 4 for crude protein, 23 for lactose and 18 for solids-not-fat. Politiek (1956) found that older cows had slightly higher protein and lower lactose than younger cows.

Griffiths and Featherstone (1957) collected monthly samples from seven herds of mostly Holstein cows during a two year period. For 336 lactation records the decline in solids-
not-fat content with increased age was approximately .1 per cent per lactation.

Comberg and Voigtländer (1958) presented lactation averages for twenty cows through the first three lactations of each. The protein content increased from 3.35 for the first lactation to 3.44 for the second then decreased to 3.38 in the third. The lactose was approximately equal in the first two lactations and decreased .13 from the second to the third.

Marckmann and Witt (1956) reported on the relation between age and protein content from more than 1,200 yearly averages of cows in two areas of Schleswig-Holstein. They observed a total decrease of .14 per cent protein from ages less than four to over ten years. The protein content was determined by formol titration.

Influence of Stage of Lactation

The reports on the influence of lactation stage on milk composition are numerous. However, except for the analysis of Waite et al. (1956), the effects of lactation stage and months of the year have not been separated completely. The rapid decrease of fat and protein content after parturition is well known. Some disagreement exists over the minimum point of the lactation curve, although most of the recent reports show this minimum to occur approximately six weeks after
calving. Apparently the lactose content decreases almost continuously over the whole lactation, although the slope of the curve has been slight in some studies. After fat, protein and solids-not-fat reach the minimum point, the increase during the remainder of the lactation appears to be nearly linear.

Waite et al. (1956) used tests from 4,988 daily composites to investigate the effects of lactation stage. Total solids, casein, total protein and solids-not-fat decreased rapidly for the first 45 days and total solids fell for another 30 days. The concentrations of these constituents increased continuously during the remainder of the lactation, rising more rapidly after about 200 days. The curve for lactose content was opposite to that for protein. The range was approximately .3 for solids-not-fat, .9 for total solids, and .7 for protein. Casein content increased less rapidly than total protein during the latter part of lactation.

Politiek (1956) derived lactation curves from tests throughout the lactation period from 111 paternal sisters. Protein and solids-not-fat reached a minimum at about the sixth week. They both increased to the end of the lactation, with protein increasing about .6 and solids-not-fat approximately .3. A corresponding decrease of .3 occurred in the lactose content during the same period.
Hansson et al. (1950) analyzed 1,306 individual samples from 67 cows. They observed the low points of total protein, casein and albumin to occur at approximately the sixth week of lactation. All three constituents increased steadily during the remainder of the lactation. Casein content increased at a slower rate than total protein. Total protein differed from the sixth to the 44th week of lactation by about .8 per cent.

Jarrige and Rossetti (1957) computed lactation curves for the protein content of 28 cows of two breeds. The minimum was reached in the second month and the increase was more rapid after the fifth month of lactation. The protein content increased approximately 1.0 from the second month until the end of the lactation.

Comberg and Voigtlander (1958) presented lactation curves calculated from more than 2,000 analyses from 149 lactations of 88 cows. Protein was lowest in the second month of lactation and during the remainder of the lactation exhibited a steady total increase of about 1.0 per cent. An approximate decrease of .3 occurred in the lactose content over the whole period. The ash content increased slightly, and at a faster rate during the last two months of the lactation.

Griffiths and Featherstone (1957) investigated the relation between stage of lactation and solids-not-fat for dif-
ferent herds. Davis et al. (1947) derived curves for differ­
ent breeds. These curves appeared to be generally the
same; however, the authors failed to separate causes of vari­
atation which could have accounted for the observed differences.

Influence of Month of Year

The influence of month of year is a compendium of the
nearly always inseparable effects of month to month changes
in temperature, nutrition and other factors including man­
agement. Until recent years, the opinion was prevalent that
nutritional levels had little effect on milk composition.
However, several recent experiments have shown that especially
the protein content can be decreased considerably by low ener­
gy diets. Burt (1957) has reviewed comprehensively the lit­
erature on this subject.

The effects of temperature on milk composition have been
studied by keeping cows in psychrometric rooms, by Regan and
Richardson (1938), and Cobble and Herman (1951). These ex­
periments established a threshold in the vicinity of 85 de­
grees Fahrenheit, above which fat content increased and
solids-not-fat and protein percentage both decreased. Cobble
and Herman also noted an increase in protein content at tem­
peratures below 30 degrees.

Waite et al. (1956) found that monthly means of the
protein and lactose contents mostly varied inversely and, thus, solids-not-fat varied less from month to month than either of its two main components. They found a sharp increase in protein content soon after the cows were first pastured in the spring. These observations were made after the influence of lactation stage was eliminated, and the authors state that seasonal variations are smaller than variation between lactation stages. Politiek (1956) also noted an increase in protein content soon after the cows were turned out to pasture.

After summarizing twelve reports on the effect of season, Herrmann (1954) concluded that the maximum solids-not-fat content usually was reached in early winter between October and January and the minimum occurred during the period of April to September. The largest difference between any two months for the simple average of these twelve studies was .19 per cent between January and July. Herrmann noted that the effect of season varied greatly in the different studies.

Reinart and Nesbitt (1956) recently reported a significant interaction between months and years for protein content from herd-milk samples taken in Canada from 12 farms every month for two years. Indeed, the month of May had the minimum protein content one year and the maximum the next. On
the other hand, the seasonal effects on protein content presented by Auriol and Mocquot (1957) for one region of France seemed very similar from year to year.

Specht et al. (1956) studied the influence of season on milk composition in two Jersey herds in Michigan. Protein content was highest from October to December and lowest during the July - September period. The difference between the two seasons was .22 per cent protein. They presented data for twelve months on protein content and for fifteen months on solids-not-fat. The one season, April to June, which was duplicated during the fifteen months contained both the largest (9.69) and smallest (9.37) contents of solids-not-fat.

Overman (1945) tabulated by months and breeds the data from 2,426 milk samples which had been collected at the Illinois station. Overman indicated that protein, fat, and solids-not-fat tend to vary together from month to month and the lactose tends to vary inversely to the other constituents. The different breeds exhibited similar trends in monthly variations. In the combined data the lowest protein content occurred in June and July and the highest from October to February. The range was .22 in protein content and .12 for solids-not-fat. The largest values for lactose were from May to July and the range was .17 per cent.

Overman et al. (1953) sampled bulk milk of the same
29 Brown Swiss herds for 12 months. The herds were located in several different states. In the averages for each month they found all the major constituents to be higher in the winter and lower during the summer months. The difference between the highest and lowest months was .27 between November and July for total protein, .11 between September and April for lactose, .35 between December and August for solids-not-fat and .67 between October and August for total solids.

After analyzing approximately 20,000 samples of herd milk, Jack et al. (1951) concluded that in California the maximum solids-not-fat content was reached in the winter and the minimum occurred during the summer. However, their curves showed considerable variability from one year to the next.

Influence of Mastitis

Mastitis, especially the more acute cases, certainly affects milk composition. Nevertheless, the practical consequences of subclinical mastitis as a cause of variation in milk composition remain to be determined. By the precise method of comparing milk composition of diseased quarters with that of normal quarters from the same cow, it has been established that mastitis causes a decrease in solids-not-fat content in the diseased quarters and this decrease is almost wholly due to decreased lactose. Casein content is also lowered, but total nitrogen remains virtually constant. The
chloride content increases, supposedly to counterbalance the decrease in lactose and to maintain the osmotic relationship between blood and milk.

Rowland and Zein-el-Dine (1938) collected samples from 247 quarters of 62 cows. The solids-not-fat content of normal samples varied from 8.24 to 10.28 while the samples from mastitic quarters ranged from 4.26 to 9.92 per cent. Non-fat solids percentage in the normal samples was higher than that in the abnormals by .92 in Shorthorns, .73 in Friesians and .44 in Ayrshires.

Richardson et al. (1950) grouped 701 samples of individual cow's milk from two breeds according to whether or not the samples came from mastitic udders. The observed differences (the normals minus the abnormals) were .43 in Holsteins and .11 for Jerseys. However, they failed to account for the fact that the normals and abnormals may not have been equally frequent for different ages, lactation stages, etc. The frequencies of abnormals were 31 per cent for Holsteins and 23 per cent for Jerseys.

Waite and Blackburn (1957) counted the total and polymorph cells in approximately 2,000 daily composites of individual Ayrshire cow's milk and 823 bulk farm milk samples. The percentage of samples containing polymorph cells increased from 14 per cent during the first lactation to approximately 60 per cent for the fifth and later lactations. The distribu-
tion of total cells followed the same general increase with age. The frequency of polymorph cells varied considerably among the different months of lactation, but the largest counts occurred during early lactation. Total cell count decreased slightly during the first two months of lactation then increased steadily until the lactation terminated. Although recognizing that cell counts are confounded with ages and lactation stages, the authors failed to separate these confounded effects in their analysis. Waite and Blackburn found that solids-not-fat content decreased about .2 as the total cell count increased to 1,000,000/ml. Approximately the same effect was found in bulk farm milk as total cell count increased.

McKenzie et al. (1958) compiled separate curves for the effect of age on solids-not-fat for milk from cows with low and high cell counts. The solids-not-fat contents for the normals and abnormals both followed the general decrease with age already noted, but solids-not-fat was slightly lower for the abnormals in all lactations except the first. They also found that low values of solids-not-fat occurred more frequently in the daily composites from individual cows with high cell counts but that this association was doubtful in bulk herd milks.
Campbell et al. (1955) investigated the repeatability of solids-not-fat content with 495 lactations of 161 cows. They found a value of .76 for repeatability, but this estimate apparently contained some breed differences. The regression of solids-not-fat on fat content for the 495 lactations was approximately .38.

Provan (1955) stated that the Milk Marketing Board obtained correlations between lactation averages of fat and solids-not-fat contents. Apparently the correlations ranged from .34 to .60 for nine herds. The average regression of solids-not-fat on fat for the nine herds was .4.

Several workers have used lactation averages to correlate fat content with protein content. Lonka (1947) found this correlation to be .60 for records of 54 West Finnish cows. From 149 lactations of 88 cows, Comberg and Voigtlander (1958) reported .45 for this correlation. Ketelaars (1956), using records from 57 daughters of two sires, observed a value of .50 for this statistic. Politiek (1957) reported correlations of .51 and .54 for 252 and 300 cows in two different areas of the Netherlands.

Johansson and Claesson (1957) used data from Gaines and Overman (1938) to calculate intra-breed correlations
between percentages of milk constituents for 305 day lactation records. With 67 degrees of freedom the correlations were .70 for fat with protein, .04 between fat and lactose and .12 between protein and lactose.

Harvey et al. (1954) sampled a herd of Holsteins and Jerseys each month for eleven months. Correlations were calculated on an inter-cow intra-breed basis. For 78 cows fat was correlated with protein .62, with solids-not-fat .51 and with total solids .90. Protein was correlated with solids-not-fat .69 and with total solids .75. The correlation between solids-not-fat and total solids was .84.

Vanschoubroek and Willems (1955) calculated the correlation between fat and protein contents from single samples of 218 cows. The samples were taken during June and July and only when each cow was in the fourth month of lactation. They state that the correlation (.29) is genetic; however, in reality it is a phenotypic correlation.

Marckmann and Witt (1956) reported .58 as the correlation coefficient between yearly averages of protein and fat contents.

One of the few reports on genetic statistics concerning milk composition was made by Johnson (1957). Johnson calculated the statistics on an intra-sire basis from lactation
averages of 76 Holstein and 70 Jersey daughter-dam pairs. Heritabilities of fat, solids-not-fat and total solids contents closely approximated .35 in both breeds. Genetic correlations for the Holsteins were .72 between fat and solids-not-fat, .85 for fat with total solids and .99 between solids-not-fat and total solids. Corresponding genetic correlations for the Jerseys were .35, .38 and .98 respectively. The phenotypic correlation between fat and solids-not-fat contents was almost exactly .37 in both breeds.

Stewart and O'Connor (1958) reported genetic correlations between fat and solids-not-fat contents for lactation averages. For 53 sire groups of 683 Friesian cows this correlation was .33 ± .16 and for 516 Ayrshires from 35 sires it was .40 ± .17.

Robertson et al. (1956) reported heritabilities, and genetic and phenotypic correlations derived from lactation records on 500 daughter-dam pairs of Ayrshires. Heritabilities of the milk constituent percentages were: .32 for fat, .53 for solids-not-fat, .48 for protein, and .36 for lactose. Genetic correlations were: .46 between fat and solids-not-fat, .48 for fat and protein, .37 for fat with lactose, .94 for solids-not-fat with protein, .67 for solids-not-fat with lactose and .41 between protein and lactose. Corresponding phenotypic correlations (for daughters only) were: .40, .42, .11, .81, .50 and -.01.
Politiek (1956) states that in his data heritabilities of fat, protein, lactose and solids-not-fat contents were approximately equal (ranged from .5 to .7). His data included progeny groups of about 50 cows each from six sires and 199 daughter-dam pairs. These estimates apparently contain some herd differences.

Considerable experimentation has been done on milk composition with identical twins in Sweden and New Zealand. For milk and fat yields, the heritabilities calculated from data on identical twins have been much higher than the same statistics derived from cows related less closely. Brumby (1958) has enumerated the several possible reasons which have been advanced to explain these discrepancies. Apparently, no general unanimity exists as to the most probable cause.

Hancock (1953) subjected identical twins to three levels of nutrition in an incomplete block experiment. The experiment was conducted during three seasons. During this period 29 sets of twins were used. Heritabilities of fat and casein contents were .95 and .94 respectively. The genetic correlation between these two constituents was .67.

Hansson (1956) summarized the research done on milk composition by himself and Bonnier. Holding fat percentage constant, they calculated the variance for protein and lactose
contents among cows with equal coefficients of relationship. When these variances were plotted for degrees of relationship ranging from zero to unity, they decreased as the relationship increased. The decrease in the variance for protein content appeared to be linear. This indicated that the genetic variance for protein content, at fixed phenotypes for fat content, was mostly additive. The decrease in the variance for lactose content was slightly curvilinear, indicating that some dominance or epistasis was involved.

To circumvent the criticism that the variance within pairs of identical twins is biased downwards because of a common environment, Hansson (1956) adjusted this variance with an estimate of the bias derived from "blind" pairs. Blind pairs are unrelated but contemporary animals subjected to similar environmental conditions as the twins. Heritabilities, after the bias supposedly was removed, were .87 for fat, .88 for protein and .62 for lactose contents.

Recent reports have indicated that the Beta-lactoglobulin fraction of the protein content can be separated into two parts, A and B. By the paper electrophoresis method Aschaffenburg and Drewry (1957) observed that the milk of individual animals contained the faster-moving A, the slower-moving B or a combination of the two. These authors postulated that
this phenomenon was controlled by a single pair of allelic genes with incomplete dominance. Data were presented from 278 animals. The observed frequencies and the frequencies expected under the hypotheses of one pair of alleles and random mating agreed remarkably well. In 24 sets of identical twins the members of each set showed perfect concordance for the electrophoretic pattern.
NATURE AND SOURCE OF DATA

The data used in this study were determinations of milk constituents in samples of milk from individual cows. Daily composites (three-day in two herds) were taken from each cow unless she was in oestrus or known to be sick. The composites were collected by personnel of Oregon State College from December, 1948 to January, 1955. Herds of the five major dairy breeds were studied as follows: one Ayrshire, two Brown Swiss, five Guernsey, seven Holstein and ten Jersey. Two of the farms had both Jerseys and Holsteins. Throughout the analyses these two were treated as four separate herds.

Each herd was under study for a total period of about 12 to 14 months. It was intended to visit each herd once per month to collect the daily composites. But, for one reason or another, this schedule varied so that the total number of visits ranged from 7 to 12 and the average interval between visits was approximately six weeks. Usually at the last time each herd was visited, information was obtained as to each cow’s sire, dam, birth date and pertinent calving dates.

After each cow was prepared for milking a sample from each of her quarters was examined for mastitis with a strip-cup and a bromthymol blotter. Then a sample from each quarter was drawn into a sterile test tube for further mastitis tests at the college laboratory. If a cow obviously had mastitis,
the milk from the affected quarters was not included in the daily composite.

At the college laboratory each sample of foremilk was analyzed for chloride content, leucocyte count, presence of long chain streptococci and by bromthymol blue. If at least one of the cow's quarters was classified as abnormal by two or more of the above tests, the corresponding daily composite was classified as abnormal.

The criterion used for separating the samples on leucocyte count was 1,000,000/ml. The bromthymol blue test color chart, Plastridge and Anderson (1933), ranges from one to eight. Colors one and two are normal, three suspicious, and four to eight abnormal. Cows in late lactation show a higher color number. A cow in late lactation with four quarters showing number three was considered normal, however three quarters showing number three and one quarter five was abnormal. In other words, the lowest color number of a particular composite was used as the standard unless all were four or above. An intra-cow standard was used also for chloride. A chloride content of .14 was considered the upper limit, however if three quarters showed .09 per cent chloride and one quarter .13 the composite was classified abnormal for chloride.

The daily composites were analyzed for total solids by
the Mojonnier procedure, fat was obtained by the Babcock test and solids-not-fat were calculated by difference. Protein content was determined by formol titration as outlined by Richardson et al. (1953).

Since the primary purpose of this investigation was to estimate the normal composition of cow's milk in Oregon, herds were sampled in the principal dairying areas of the state. Herds were selected within these areas according to breed, willingness of the owner to cooperate, and availability of records pertinent to this study. The management of the herds can be described roughly as varying from the better managed herds on DHIA to herds kept under "AR conditions". The herds within each breed also can be classified approximately as ranging from "high grades" to "elite purebreds".

The location of the herds in the different areas of the state is shown in Figure 1. As will be the custom in the rest of this manuscript, the letter within each circle refers to the breed and the number designates a particular herd. The dotted lines enclose areas with similar climates and geography. Table 1, statistics from Highsmith (1958), gives the pertinent data corresponding to the marked areas in Figure 1. The enclosed areas in Figure 1 are not intended to be topographically accurate.
Figure 1. Map of Oregon showing location of herds studied
Table 1. Climatic statistics pertaining to the areas in Figure 1

<table>
<thead>
<tr>
<th>Approximate elevation, (feet)</th>
<th>I</th>
<th>II</th>
<th>Area</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yearly precipitation, (inches)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>January average maximum</td>
<td>50-55</td>
<td>40-45</td>
<td>35-40</td>
<td>40-45</td>
<td>30-35</td>
<td></td>
</tr>
<tr>
<td>temperature (degrees Fahrenheit)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>January average minimum</td>
<td>30-35</td>
<td>30-35</td>
<td>15-25</td>
<td>30-35</td>
<td>15-20</td>
<td></td>
</tr>
<tr>
<td>temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>July average maximum</td>
<td>65-70</td>
<td>75-85</td>
<td>80-85</td>
<td>90-95</td>
<td>90-95</td>
<td></td>
</tr>
<tr>
<td>temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>July average minimum</td>
<td>50-55</td>
<td>45-55</td>
<td>40-45</td>
<td>45-55</td>
<td>55-60</td>
<td></td>
</tr>
<tr>
<td>temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average date of last 32 degree frost in spring</td>
<td>Before Mar. 31</td>
<td>Before Apr. 15</td>
<td>Before May 30</td>
<td>Before Apr. 30</td>
<td>Before May 15</td>
<td></td>
</tr>
<tr>
<td>Average date of first 32 degree frost in fall</td>
<td>After Nov. 15</td>
<td>Before Oct. 30</td>
<td>Before Aug. 15</td>
<td>Before Oct. 15</td>
<td>Before Sept. 30</td>
<td></td>
</tr>
</tbody>
</table>
As Table 1 shows, the herds were kept under a very wide range of climatic conditions. The inches of precipitation shown in Table 1 do not indicate the amount of pasture available because 18 of the 23 farms had irrigation. Of the five farms without irrigated pasture, two were in area II and three in area I.
METHODS OF ANALYSIS AND RESULTS

Transformations

Having data from five breeds is advantageous in one sense because it provides a wider inductive basis for any general conclusions one wishes to make, but it does present some problems. If the different analyses to be performed were done separately for each breed the computations would have been prohibitive. Also the volume of data appeared barely sufficient to evaluate reasonably the answers wanted. This seemed especially so for estimates of genetic parameters. Even if the various estimates were calculated for each breed, it seemed improbable that a sound basis would exist for judging whether differences were genuinely due to breeds or could have resulted easily from sampling errors. This is because formulae for standard errors are not yet generally available for variance components computed from non-orthogonal data or from data which have been adjusted with correction factors estimated from the same data.

Estimation of components of variance and covariance is fundamental in this type of study. Since it was desired to pool the estimates over breeds, the question arose as to how much the variances within herds differed from breed to breed. This question is important because, if heterogeneity of variance exists, the contribution of the sample from each breed
to the pooled estimate would not be entirely a function of the degrees of freedom in each sample as it should be, but would be a combination of the degrees of freedom, and the size of the sample variance.

For these reasons it was decided to calculate the total variance within herds separately for each breed and, if it then appeared necessary, to seek suitable transformations. Of course, it already is well-known that the fat content is more variable in higher testing than in lower testing breeds. The mean intra-herd variance is plotted against the breed mean for each breed, in Figures 2 to 5. To calculate these variances, 5,621 observations were used. The degrees of freedom for each breed ranged from 184 for Ayrshires to 1,793 for Holsteins.

Figures 2 to 5 show definite differences among breeds in their intra-herd variances. Considering the numbers involved, tests of significance hardly seemed necessary to demonstrate that these differences were real. The differences are much larger for fat and total solids than for protein and solids-not-fat. As will be shown later, many sources contribute to the total variance within each herd, but reasons for these sources differing much from breed to breed are not apparent.

The classical discussion on transformations was given
Figure 2. Relation between mean and variance for fat content

Figure 3. Relation between mean and variance for protein content
Figure 4. Relation between mean and variance for solids-not-fat content

Figure 5. Relation between mean and variance for total solids content
VARIANCE WITHIN HERDS

120
50
130
H
A
B

TOTAL SOLIDS CONTENT

130
140
150

85
8.9
9.3
9.7
H
A
B
G
J

SOLIDS-NOT-FAT CONTENT

8.5
8.8
9.8

VARIANCE WITHIN HERDS

13
18
23
by Bartlett (1947). He stated that if the variance can be described as a function of the mean, say, \( \sigma^2 = g(\mu) \) the approximate transformation to correct heterogeneity is obtained by evaluating the indefinite integral,

\[
\int \frac{dX}{\sqrt{g(X)}} = f(X).
\]

One difficulty here is that this transformation only corrects heterogeneity for one classification of means. If a model with many effects is appropriate the relation between the mean and the variance may be different depending on the effect involved.

Sometimes a priori knowledge would exist of the relation between the mean and the variance. For instance, if the data followed the Poisson distribution, the variance equals the mean and the appropriate transformation is \( \sqrt{X} \).

No a priori reason seemed obvious for assuming any particular relationship between the means and variances in these data. Supposedly, the reason the variances generally decreased with the means was because as the mean decreased the variance was being damped down by a lower limit beyond which it was difficult for the constituents to become smaller. This situation would be analogous to the distribution of the sample correlation coefficient where, as the true value approaches minus unity, the distribution becomes more extremely skewed.
with the long tail toward zero. The distribution for fat content, and to a lesser extent that for total solids, did show this general type of skew distribution when the cumulative frequencies were plotted on probability paper. However, this wasn't the case for protein and solids-not-fat content; since those had longer tails in both directions than would be expected if they were exactly normally distributed. Tocher (1925) postulated a Pearson type IV curve for solids-not-fat and these data seem to agree with this. Gowen (1924) states that in Holsteins the lactation averages for both fat and solids-not-fat have type IV distributions.

It was decided that the best procedure would be to estimate the relationships between the means and variances from the data in Figures 2 to 5. Except for solids-not-fat, linear regression appeared to describe $V(X)$ as a function of $\bar{X}$ about as accurately as any reasonably simple function. Since the differences among the variances were not nearly as large for solids-not-fat as for fat and total solids and, thus, a transformation not as important, a linear regression also was used for solids-not-fat.

When $\mu$ is put approximately equal to $g(\bar{X})$ and the integral evaluated, $\int \frac{dX}{\sqrt{a + bX}} = \frac{2}{b} \sqrt{a + b\bar{X}}$. Because the variance is unchanged when the observations are multiplied by a constant, $\sqrt{X + \frac{a}{b}}$ is the estimated transformation. Since
the \( \frac{a}{b} \) values for these data were all negative and, except for solids-not-fat, they were of the size that the quantity \( (X - \frac{a}{b}) \) could be negative, these values were modified slightly to ensure that \( (X - \frac{a}{b}) \) would be positive.

The variances for the transformed data are given in Table 2. In the remainder of this dissertation it will often be convenient to use symbols for the various milk constituents so let \( F = \) fat, \( P = \) protein, \( S = \) solids-not-fat, and \( T = \) total solids. The transformations used were \( \sqrt{F-1.9}, \sqrt{P-1.7}, \sqrt{S-4.0} \) and \( \sqrt{T-9.5} \).

Table 2 shows that the transformations removed most of the heterogeneity, but were not completely effective. The test for homogeneity of variance proposed by Hartley (1950) shows the variances of the transformed data to be significantly different at \( P \leq .05 \) for every constituent. When the variances for the transformed data were compared with Figures 2 to 5, the largest discrepancies seemed to occur when the original variance for that breed deviated widely from the straight line. For instance, the transformation for total solids was very satisfactory except for Brown Swiss, and Figure 5 showed that Brown Swiss deviated most from the linear relationship. Consequently, it was considered impractical to seek better transformations.
Table 2. Pooled intra-herd variances for each breed

<table>
<thead>
<tr>
<th>Breed</th>
<th>D.f.</th>
<th>V(F)</th>
<th>V(VF-1.9)</th>
<th>V(P)</th>
<th>V(√P-1.7)</th>
<th>V(S)</th>
<th>V(√S-4)</th>
<th>Total solids %</th>
<th>V(T)</th>
<th>V(√T-9.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayrshire</td>
<td>184</td>
<td>.32</td>
<td>.041</td>
<td>.12</td>
<td>.015</td>
<td>.15</td>
<td>.0080</td>
<td>.74</td>
<td>.060</td>
<td></td>
</tr>
<tr>
<td>Brown Swiss</td>
<td>614</td>
<td>.27</td>
<td>.036</td>
<td>.11</td>
<td>.019</td>
<td>.14</td>
<td>.0067</td>
<td>.53</td>
<td>.035</td>
<td></td>
</tr>
<tr>
<td>Guernsey</td>
<td>1230</td>
<td>.63</td>
<td>.052</td>
<td>.13</td>
<td>.019</td>
<td>.20</td>
<td>.0092</td>
<td>1.08</td>
<td>.056</td>
<td></td>
</tr>
<tr>
<td>Holstein</td>
<td>1793</td>
<td>.29</td>
<td>.047</td>
<td>.08</td>
<td>.014</td>
<td>.20</td>
<td>.0107</td>
<td>.65</td>
<td>.064</td>
<td></td>
</tr>
<tr>
<td>Jersey</td>
<td>1774</td>
<td>.73</td>
<td>.050</td>
<td>.14</td>
<td>.017</td>
<td>.21</td>
<td>.0096</td>
<td>1.29</td>
<td>.057</td>
<td></td>
</tr>
</tbody>
</table>
Factors Affecting the Composition of Individual Samples of Cow's Milk

The literature on different environmental and physiological factors affecting the variation of milk constituents is voluminous. However, except perhaps for fat content, no general unanimity on the relative importance of these factors is apparent. Their relative magnitudes are, of course, not expected to be exactly the same in different sets of data.

It was considered appropriate to investigate, insofar as practical with the data at hand, the relative magnitudes of some of these factors. This information was needed not only for its own value but also for formulating a mathematical model for the genetic analysis.

The different factors included in such an analysis are to some extent dictated by the data. For instance, it would have been desirable to include here the cow's stage of gestation when the sample was taken. This was not possible because this information had not been collected for enough cows. Apparently the factors most likely to be important were the cow herself, her age, her stage of lactation, the herd and breed to which she belonged, whether or not her foremilk had been classified abnormal, and the month of the year. In addition, at least some of the interactions among these factors appeared to merit attention.
These data are highly non-orthogonal, i.e. the effects of the different factors are superimposed or confounded with one another. For instance, each herd was not studied in each different month. This non-orthogonality made choosing a method to evaluate the relative magnitudes of the different factors a perplexing problem. The standard method for uncorrelated random variables is to calculate the mean squares as if the data were orthogonal, equate these mean squares to their expected values and simultaneously solve the resulting equations for the estimates of variance components. Correlations among the factors in the model can bias seriously the variance components estimated by this method, but no reasons for expecting important correlations among these factors were apparent.

One difficulty with these data is that some of the factors such as age of cow, stage of lactation, etc., may be regarded more appropriately as "fixed" than random variables. As Henderson (1953) stated, if some of the factors are "fixed" this method yields biased estimates. Unfortunately, any general practical consequences of this bias remain to be determined.

Seemingly, the only alternative was to conduct a complete least squares analysis Yates (1934). If the interaction terms were included in the model this would involve solving hundreds
of simultaneous equations. A joint significance test can be made for all the interactions, Wilks (1938), but it was feared that the power of this test would be small for any particular interaction. Moreover, the object was to estimate variance components rather than make significance tests. This is because we already are confident that, at least for the main effects, each variance component is not zero. The information wished is their relative magnitudes.

It was concluded that the best practical way the factors could be considered simultaneously and at the same time obtain the wanted information was to proceed as if all factors were uncorrelated and random, equate the mean squares to their expectations under these assumptions and solve for the variance components. The theory and mechanics of this method are amply discussed by Henderson (1953), Kempthorne (1957) and others.

For this part of the analysis, 4,462 observations were available for each milk constituent. Two herds, HO and JO, were discarded because most of the cows were grades and some pertinent information such as birth dates was absent. Age was calculated as the age of cow at date of calving. Cows ten years old and older were grouped together. Stage of lactation was coded to the nearest month. Tests during the first ten days and after the tenth month were discarded. The analysis was conducted on both the transformed and original data.
It would have been desirable to include all of the two-factor interactions in the model, but this seemed impractical because of computational labor. Among all the main effects, abnormality of the udder seemed least likely to enter into interactions. This may or may not have been a tenable assumption.

The following linear model was used to describe the data.

\[ X_{ijkpqrs} = \lambda + h_i + m_j + l_k + a_p + u_q + (hm)_{ij} + (h1)_{ik} \\
+ (ha)_{ip} + (ml)_{jk} + (ma)_{jp} + (1a)_{kr} + c_{ir} \\
+ e_{ijkpqrs} \]

where: \( X_{ijkpqrs} \) is the measure of a milk constituent in a daily composite from the \( r \)th cow with the \( q \)th udder condition at the \( p \)th age in the \( k \)th stage of lactation during the \( j \)th month of the year in the \( i \)th herd.

The general mean is \( \lambda \) and since all of the effects included in the model will be estimated and discussed as deviations it will not be considered further.

Each term in the model measures the sum of the individual effects which make that particular classification's total effect on a pertinent milk constituent differ from the average of all the members of that classification. For instance, the herd effect, \( h_i \), measures the totality of different effects
which make the average of a particular milk constituent in the \(i^{th}\) herd differ from the mean of all herds. An interaction term, say, \((hm)_{ij}\) is defined as the sum of the effects which make the average of the samples for the \(ij^{th}\) herd-month different from the average effects of the \(i^{th}\) herd and \(j^{th}\) month.

The assumptions ordinarily made in using this method are uncorrelated random variables with zero means and variances; 
\[\sigma_H^2, \sigma_M^2, \ldots, \sigma_E^2.\] However, as mentioned before, the assumption of random variables is not completely fulfilled for those factors exhibiting trends.

The results from this analysis are given in Tables 3 to 6. Table 3 shows the coefficients of the components in the expected mean squares. The underlined values in Table 3 indicate those cells where expectations would appear if, in fact, the data were orthogonal and the variables were random; all other cells would be zero. The negative expectations for the error mean square arise because not all the possible interactions were included in the model. Table 3 illustrates the extreme non-orthogonality between some of the variables. For example, the expectations of 12.85 V(\(U\)) in the mean square for ages and 35.19 V(A) in the mean square for udder conditions result from the frequency of damaged udder tissue and mastitis being much higher in older than younger cows. Be
<table>
<thead>
<tr>
<th>Source of variation</th>
<th>D.f.</th>
<th>$\sigma^2_H$</th>
<th>$\sigma^2_M$</th>
<th>$\sigma^2_L$</th>
<th>$\sigma^2_A$</th>
<th>$\sigma^2_U$</th>
<th>$\sigma^2_{HM}$</th>
<th>$\sigma^2_{HL}$</th>
<th>$\sigma^2_{HA}$</th>
<th>$\sigma^2_{NL}$</th>
<th>$\sigma^2_{MA}$</th>
<th>$\sigma^2_{IA}$</th>
<th>$\sigma^2_C$</th>
<th>$\sigma^2_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herds, (H)</td>
<td>22</td>
<td>193.00</td>
<td>7.82</td>
<td>.63</td>
<td>5.17</td>
<td>9.66</td>
<td>24.58</td>
<td>20.41</td>
<td>34.51</td>
<td>2.28</td>
<td>2.37</td>
<td>1.17</td>
<td>6.34</td>
<td>1.00</td>
</tr>
<tr>
<td>Months, (M)</td>
<td>11</td>
<td>9.14</td>
<td>369.64</td>
<td>2.60</td>
<td>.70</td>
<td>1.25</td>
<td>26.34</td>
<td>2.63</td>
<td>2.13</td>
<td>40.30</td>
<td>56.17</td>
<td>1.27</td>
<td>5.59</td>
<td>1.00</td>
</tr>
<tr>
<td>Lactation stages, (L)</td>
<td>9</td>
<td>.72</td>
<td>2.74</td>
<td>444.90</td>
<td>.40</td>
<td>2.65</td>
<td>1.82</td>
<td>22.21</td>
<td>.63</td>
<td>41.92</td>
<td>1.29</td>
<td>67.18</td>
<td>.48</td>
<td>1.00</td>
</tr>
<tr>
<td>Ages, (A)</td>
<td>9</td>
<td>7.22</td>
<td>.74</td>
<td>454.32</td>
<td>12.85</td>
<td>1.33</td>
<td>1.27</td>
<td>27.04</td>
<td>1.05</td>
<td>38.07</td>
<td>63.66</td>
<td>5.08</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Udder conditions, (U)</td>
<td>1</td>
<td>26.49</td>
<td>2.33</td>
<td>5.37</td>
<td>35.19</td>
<td>1789.91</td>
<td>4.23</td>
<td>3.74</td>
<td>7.88</td>
<td>1.83</td>
<td>4.31</td>
<td>5.50</td>
<td>3.15</td>
<td>1.00</td>
</tr>
<tr>
<td>Herds x months, (HM)</td>
<td>171</td>
<td>-.59</td>
<td>-1.01</td>
<td>1.59</td>
<td>.49</td>
<td>.35</td>
<td>21.08</td>
<td>1.59</td>
<td>.40</td>
<td>1.38</td>
<td>.92</td>
<td>1.13</td>
<td>.36</td>
<td>1.00</td>
</tr>
<tr>
<td>Herds x lactation stages, (HL)</td>
<td>198</td>
<td>-.03</td>
<td>1.51</td>
<td>-.07</td>
<td>.52</td>
<td>.32</td>
<td>1.55</td>
<td>19.14</td>
<td>.51</td>
<td>2.13</td>
<td>1.10</td>
<td>1.03</td>
<td>.50</td>
<td>1.00</td>
</tr>
<tr>
<td>Herds x ages, (HA)</td>
<td>166</td>
<td>-.39</td>
<td>.49</td>
<td>.60</td>
<td>-.77</td>
<td>1.13</td>
<td>.46</td>
<td>.55</td>
<td>20.61</td>
<td>.95</td>
<td>1.21</td>
<td>.53</td>
<td>4.50</td>
<td>1.00</td>
</tr>
<tr>
<td>Months x lactation stages (HL)</td>
<td>99</td>
<td>1.64</td>
<td>-.25</td>
<td>-.29</td>
<td>1.01</td>
<td>.51</td>
<td>1.54</td>
<td>2.36</td>
<td>1.17</td>
<td>36.34</td>
<td>.92</td>
<td>.94</td>
<td>1.06</td>
<td>1.00</td>
</tr>
<tr>
<td>Months x ages, (MA)</td>
<td>99</td>
<td>.68</td>
<td>-.07</td>
<td>1.01</td>
<td>-.08</td>
<td>.44</td>
<td>1.22</td>
<td>1.05</td>
<td>1.46</td>
<td>.96</td>
<td>34.76</td>
<td>1.16</td>
<td>.65</td>
<td>1.00</td>
</tr>
<tr>
<td>Lactation stages x ages, (IA)</td>
<td>81</td>
<td>.58</td>
<td>1.05</td>
<td>-.05</td>
<td>-.05</td>
<td>.43</td>
<td>1.04</td>
<td>1.24</td>
<td>.59</td>
<td>1.01</td>
<td>1.21</td>
<td>41.92</td>
<td>.71</td>
<td>1.00</td>
</tr>
<tr>
<td>Cows within herds, (C)</td>
<td>793</td>
<td>.00</td>
<td>.39</td>
<td>.52</td>
<td>3.59</td>
<td>.86</td>
<td>.39</td>
<td>.52</td>
<td>3.59</td>
<td>.93</td>
<td>.90</td>
<td>.94</td>
<td>5.44</td>
<td>1.00</td>
</tr>
<tr>
<td>Error (E)</td>
<td>2802</td>
<td>-.11</td>
<td>-.28</td>
<td>-.32</td>
<td>-1.13</td>
<td>-.53</td>
<td>-.39</td>
<td>-.45</td>
<td>-1.20</td>
<td>-.64</td>
<td>-.55</td>
<td>-.53</td>
<td>-.43</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 4. Mean squares for analysis of variation of factors affecting the composition of samples of individual cow's milk

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>D.f.</th>
<th>Fat</th>
<th>((F-1.9)^{\frac{1}{2}})</th>
<th>Protein</th>
<th>((F-1.7)^{\frac{1}{2}})</th>
<th>Solids-not-fat ((S-4.0)^{\frac{1}{2}})</th>
<th>Total solids ((T-9.5)^{\frac{1}{2}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herds</td>
<td>22</td>
<td>157.33</td>
<td>15.000</td>
<td>23.77</td>
<td>3.371</td>
<td>35.86</td>
<td>1.757</td>
</tr>
<tr>
<td>Months</td>
<td>11</td>
<td>31.65</td>
<td>2.767</td>
<td>2.22</td>
<td>.313</td>
<td>5.64</td>
<td>.263</td>
</tr>
<tr>
<td>Lactation stages</td>
<td>9</td>
<td>28.57</td>
<td>2.375</td>
<td>8.86</td>
<td>1.221</td>
<td>8.35</td>
<td>.393</td>
</tr>
<tr>
<td>Ages</td>
<td>9</td>
<td>11.17</td>
<td>1.184</td>
<td>1.86</td>
<td>.267</td>
<td>7.11</td>
<td>.354</td>
</tr>
<tr>
<td>Udder conditions</td>
<td>1</td>
<td>15.90</td>
<td>1.309</td>
<td>.05</td>
<td>.002</td>
<td>7.78</td>
<td>1.48</td>
</tr>
<tr>
<td>Herds x months</td>
<td>171</td>
<td>.60</td>
<td>.035</td>
<td>.46</td>
<td>.062</td>
<td>.43</td>
<td>.020</td>
</tr>
<tr>
<td>Herds x lactation stages</td>
<td>198</td>
<td>.94</td>
<td>.071</td>
<td>.10</td>
<td>.013</td>
<td>.20</td>
<td>.010</td>
</tr>
<tr>
<td>Herds x ages</td>
<td>166</td>
<td>.90</td>
<td>.082</td>
<td>.23</td>
<td>.033</td>
<td>.39</td>
<td>.020</td>
</tr>
<tr>
<td>Months x lactation stages</td>
<td>99</td>
<td>.93</td>
<td>.086</td>
<td>.19</td>
<td>.026</td>
<td>.32</td>
<td>.016</td>
</tr>
<tr>
<td>Months x ages</td>
<td>99</td>
<td>.85</td>
<td>.084</td>
<td>.21</td>
<td>.029</td>
<td>.35</td>
<td>.018</td>
</tr>
<tr>
<td>Lactation stages x ages</td>
<td>81</td>
<td>1.14</td>
<td>.115</td>
<td>.26</td>
<td>.036</td>
<td>.37</td>
<td>.018</td>
</tr>
<tr>
<td>Cows within herds</td>
<td>793</td>
<td>1.23</td>
<td>.113</td>
<td>.25</td>
<td>.034</td>
<td>.48</td>
<td>.024</td>
</tr>
<tr>
<td>Error</td>
<td>2802</td>
<td>.08</td>
<td>.004</td>
<td>.01</td>
<td>.001</td>
<td>.02</td>
<td>.001</td>
</tr>
</tbody>
</table>
Table 5. Estimated components of variance for factors affecting the composition of samples of individual cow's milk

<table>
<thead>
<tr>
<th>Source of component</th>
<th>D.f.</th>
<th>Fat</th>
<th>((F-1.9)^{1/3})</th>
<th>Protein</th>
<th>((F-1.7)^{1/3})</th>
<th>Solids-not-fat</th>
<th>Total solids</th>
<th>((T-9.5)^{1/3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herds, (V(H))</td>
<td>22</td>
<td>.796</td>
<td>.0761</td>
<td>1.183</td>
<td>.01681</td>
<td>.1789</td>
<td>.008766</td>
<td>.1668</td>
</tr>
<tr>
<td>Months, (V(M))</td>
<td>11</td>
<td>.065</td>
<td>.0056</td>
<td>.0015</td>
<td>.00022</td>
<td>.0091</td>
<td>.000411</td>
<td>.126</td>
</tr>
<tr>
<td>Lactation stages, (V(L))</td>
<td>9</td>
<td>.061</td>
<td>.0051</td>
<td>.0194</td>
<td>.00267</td>
<td>.0179</td>
<td>.00084</td>
<td>.146</td>
</tr>
<tr>
<td>Ages, (V(A))</td>
<td>9</td>
<td>.009</td>
<td>.0011</td>
<td>.0013</td>
<td>.00019</td>
<td>.0120</td>
<td>.00059</td>
<td>.047</td>
</tr>
<tr>
<td>Udder conditions, (V(U))</td>
<td>1</td>
<td>-.004</td>
<td>-.0005</td>
<td>-.0020</td>
<td>-.00028</td>
<td>.0012</td>
<td>.00008</td>
<td>-.026</td>
</tr>
<tr>
<td>Herds x months, (V(HM))</td>
<td>171</td>
<td>.037</td>
<td>.0025</td>
<td>.0213</td>
<td>.00291</td>
<td>.0196</td>
<td>.00094</td>
<td>.059</td>
</tr>
<tr>
<td>Herds x lactation stages, (V(HL))</td>
<td>198</td>
<td>.032</td>
<td>.0021</td>
<td>.0012</td>
<td>.00007</td>
<td>.0034</td>
<td>.00014</td>
<td>.043</td>
</tr>
<tr>
<td>Herds x ages, (V(HA))</td>
<td>166</td>
<td>.010</td>
<td>.0010</td>
<td>.0040</td>
<td>.00062</td>
<td>.0052</td>
<td>.00030</td>
<td>.015</td>
</tr>
<tr>
<td>Months x lactation stages, (V(ML))</td>
<td>99</td>
<td>-.023</td>
<td>-.0022</td>
<td>-.0030</td>
<td>-.00045</td>
<td>-.0043</td>
<td>-.0021</td>
<td>-.048</td>
</tr>
<tr>
<td>Months x ages, (V(MA))</td>
<td>99</td>
<td>-.003</td>
<td>-.0001</td>
<td>.0007</td>
<td>.00008</td>
<td>.0021</td>
<td>.00011</td>
<td>.001</td>
</tr>
<tr>
<td>Lactation stages x ages, (V(IA))</td>
<td>81</td>
<td>.008</td>
<td>.0008</td>
<td>.0026</td>
<td>.0037</td>
<td>.0029</td>
<td>.00015</td>
<td>.019</td>
</tr>
<tr>
<td>Cows within herds, (V(C))</td>
<td>793</td>
<td>.171</td>
<td>.0154</td>
<td>.0320</td>
<td>.00424</td>
<td>.0607</td>
<td>.00295</td>
<td>.360</td>
</tr>
<tr>
<td>Error, (V(E))</td>
<td>2802</td>
<td>.158</td>
<td>.0167</td>
<td>.0345</td>
<td>.00526</td>
<td>.0651</td>
<td>.00323</td>
<td>.174</td>
</tr>
</tbody>
</table>
Table 6. Estimated components of variance for factors affecting the composition of samples of individual cow's milk, expressed as the percentage of the sum of the positive components for each constituent

<table>
<thead>
<tr>
<th>Source of component</th>
<th>D.f.</th>
<th>Fat</th>
<th>Relative magnitudes of components of variance</th>
<th>Total solids (T-9.5)%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(P-1.9)% Protein (P-1.7)% Solids-not-fat (S-4.0)% solids (T-9.5)%</td>
<td></td>
</tr>
<tr>
<td>Herds, V(H)</td>
<td>22</td>
<td>59.1</td>
<td>60.2 50.0 50.3 47.4 47.4 62.8 61.3</td>
<td></td>
</tr>
<tr>
<td>Months, V(M)</td>
<td>11</td>
<td>4.8</td>
<td>4.4 4.0 8.0 4.7 4.5 5.5 4.8</td>
<td></td>
</tr>
<tr>
<td>Lactation stages, V(L)</td>
<td>9</td>
<td>4.5</td>
<td>4.0 8.2 8.0 4.7 4.5 5.5 4.8</td>
<td></td>
</tr>
<tr>
<td>Ages, V(A)</td>
<td>9</td>
<td>.7</td>
<td>.9 8.0 8.0 3.2 3.2 1.8 2.0</td>
<td></td>
</tr>
<tr>
<td>Udder conditions, V(U)</td>
<td>1</td>
<td>-</td>
<td>- 4.0 - - .3 .4 -</td>
<td></td>
</tr>
<tr>
<td>Herds x months V(HM)</td>
<td>171</td>
<td>2.7</td>
<td>2.0 9.0 8.7 5.2 5.1 2.2 1.7</td>
<td></td>
</tr>
<tr>
<td>Herds x lactation stages, V(HL)</td>
<td>198</td>
<td>2.4</td>
<td>1.7 5.0 2.0 .9 .8 1.6 1.1</td>
<td></td>
</tr>
<tr>
<td>Herds x ages, V(HA)</td>
<td>166</td>
<td>.8</td>
<td>.8 1.7 1.9 1.4 1.6 .6 1.2</td>
<td></td>
</tr>
<tr>
<td>Months x lactation stages, V(ML)</td>
<td>99</td>
<td>-</td>
<td>- - - - - - - -</td>
<td></td>
</tr>
<tr>
<td>Months x ages, V(MA)</td>
<td>99</td>
<td>-</td>
<td>- .3 .2 .6 .6 .0 .2</td>
<td></td>
</tr>
<tr>
<td>Lactation stages x ages, V(IA)</td>
<td>81</td>
<td>.6</td>
<td>.6 1.1 1.1 .8 .8 .7 .7</td>
<td></td>
</tr>
<tr>
<td>Cows within herds, V(C)</td>
<td>793</td>
<td>12.7</td>
<td>12.2 13.5 12.7 16.0 15.9 13.5 12.9</td>
<td></td>
</tr>
<tr>
<td>Error, V(E)</td>
<td>2802</td>
<td>11.7</td>
<td>13.2 14.6 15.7 17.2 17.5 6.6 8.6</td>
<td></td>
</tr>
</tbody>
</table>
it noticed also that differences between herds are highly
confounded with udder conditions and differences between cows
are highly confounded with both ages and herd times age in-
teraction.

The computed mean squares are shown in Table 4. The
negative mean squares for error again illustrate the non-
orthogonality. Since the error is computed by subtraction,
its negativity results from subtracting each of the other
sources of variation from the total variation many more times
than they actually contributed to that total.

The usual explanation given for negative estimates of
variance components is that the true component of variance is
very near zero and the negative estimate resulted from chance
in sampling. Since the negative estimates in Table 5 are
near zero, sampling error certainly could be a tenable hy-
pothesis in this case. Negative estimates, especially for
interactions, can also result from other causes and among
these would be correlations between the main effects.

Table 5 gives the estimated variance components derived
by equating the mean squares of Table 4 to their expectations
in Table 3. It is expected that each one of these components
is freed of the other sources of variation with which it was
confounded. For example, the variance component for the dif-
ference between udder conditions is the variance expected
from this difference after allowance has been made for the fact that the abnormals were much more frequent in older than younger cows, etc.

The relative magnitudes of the components in Table 5 are shown in Table 6. The total variance used to calculate the percentages in Table 6 is not the total found in the original population. In the original population many tests would come from the same breed, age, month, etc. The total variance here is the variance expected between two random tests differing in every one of the factors in the model. Even a cursory examination of Table 6 shows that the relative magnitudes of the variance components for the original and transformed data are strikingly similar for each constituent. This appears to result from one of three causes; (i) the variance components for each of these factors are nearly alike within breeds so it does not matter much how the individual breeds are weighted in the pooled estimate, (ii) the need for transformations was overestimated or (iii) the transformations were not successful enough in removing the heterogeneity. Of these three possible causes, either (i) or (ii) or a combination of the two appears more likely than (iii).

The percentage given for each component in Table 6 is not expected to be that which would be found in data which, for instance, all came from the same herd. Since differences
between herds are such a large part of the total variation, the other sources, except interactions with herds, would be a correspondingly larger part of the total variance within herds.

The variance component between herds, $V(h)$, measures the extent to which the average of a particular milk constituent for each herd differed from the corresponding averages of the other herds. This was by far the largest source of variation, ranging from 47.3 per cent for solids-not-fat to 62.8 per cent for total solids. These differences between herds result from both genetic and environmental effects. Since breeds were not included in the model, differences between herds also contain differences between breeds. Differences between the average genetic composition of breeds and between herds within breeds comprise the genetic differences between herds. The environmental contributions to the differences between herds come from many different sources. Among these would be differences between climatic conditions of the areas where the herds are located. Because each herd was tested for only about one year, herd differences would also contain some differences between years. Finally, each herd has its own feeding and management peculiarities.

The component for months, $V(M)$, measures the difference between the average of each month and the average of every other month. This component appeared not important for pro-
tein, only moderately so for solids-not-fat and was fourth or fifth largest in fat and total solids. Differences between months contain month to month differences in the effects of feeding, management and climatic conditions, insofar as these effects are general over the whole material.

The extent to which differences between stages of lactation are a large part of the total is measured by \( V(L) \). Stage of gestation was not included in the model so some of the effects of this are also contained in \( V(L) \). This component appeared important in all of the milk constituents; about 8 per cent of the total variance for protein and 5 per cent for the others.

The often mentioned effects of stage of lactation and month of year being highly confounded are supposedly separated in \( V(M) \) and \( V(L) \). Waite et al. (1956) also separated the effects of stage of lactation and month of year. They found differences between months of the year for the same milk constituents that are measured here. Of course, equality of differences between months need not be expected in Scotland and Oregon.

The component, \( V(A) \), for ages was very small for protein (.7 per cent) and fat (.5 per cent), but was of similar magnitude to \( V(M) \) and \( V(L) \) for solids-not-fat and intermediate for total solids. Differences between ages probably are
caused mostly by the physiological effects of aging. Although
differences in the effects of age might be caused by selec-
tion, these differences are likely to have been small because,
except for fat content, the farmers had no measure of the
milk constituents for making selections. Hence, any age dif-
fferences in the constituents other than fat caused by selec-
tion would have to result either from natural selection or
through correlated responses with other characters for which
the farmers were selecting.

The component, $V(U)$, for abnormal versus normal udder
conditions measures the extent to which the average of the
constituents differed more between abnormal and normal than
would be expected from the effects of $V(H)$, $V(N)$, ... $V(C)$,
and $V(E)$ on these differences. The component, $V(U)$ was prac-
tically zero for each constituent. The abnormals included
samples from cows with active cases of mastitis, although
most of the really acute cases had been eliminated at the
farm. Also in the abnormals were samples from cows which
had recovered from mastitis but retained some damaged udder
tissue. The accepted hypothesis that abnormal udder tissue
and mastitis has its largest effect on lactose is indicated
by the fact that $V(U)$ was larger for solids-not-fat than for
protein although it was extremely small even there.

The herd-month interaction, $V(HM)$, measures the extent
the average of the samples is consistently higher or lower for one month in all herds than that average is for the other months. This component is larger for both protein (9 per cent) and solids-not-fat (5 per cent) than any of the other components of variation except herds, cows and error. For fat and total solids $V(HM)$ was larger than any of the other interaction terms. The possible causes of this interaction will be discussed later in more detail. The magnitude of $V(HM)$ for protein and solids-not-fat indicates the reason for the considerable disagreement on the effect of months in some of the literature, see Herrmann (1954), and McDowall (1936).

The herd by lactation stage interaction, $V(HL)$, measures the extent to which the differences between lactation stages are consistent from herd to herd. This component was small in these data, ranging from less than one per cent for protein and solids-not-fat to about two per cent for fat and total solids.

The component for herd-age interaction, $V(HA)$, measures the extent to which the same age correction factors for milk constituents would be applicable for all the herds. If differences between lactations by the same cow had been included in the model, $V(HA)$ probably would have been even smaller. This is because differences between lactations of the same
cow are almost completely confounded with differences between the two ages at which that cow was tested and, as will be shown later, the variance between lactations within cows is a much larger part of the total variation than herd by age interaction. As far as the variation among individual samples of cow's milk is concerned, the herd-age interaction appeared unimportant.

The component for month by stage of lactation was negative for each constituent. This could be caused by a correlation between months and stages of lactation because these factors are highly confounded on a within cow basis. If the cows tended to calve in the same season they would be tested mostly during the same months and lactation stages. This would cause the frequencies of the observations to cluster around the diagonal in a two-way table of months and stages of lactation. Most of the herd owners tried to maintain an equal distribution of fresh cows throughout the year because, except for those in the coastal region, they were selling milk to be bottled. However, those herd owners in the coastal region of the state (area I) sold milk to cheese factories and nearly all of their cows freshened during February and March. Thus, a slight correlation between months of the year and stages of lactation is expected. Probably, even if the correlation were zero, $V(ML)$ would not be a large part of the total variation.
The month-age interaction, $V(MA)$, was near zero in each case and no reason is obvious for thinking it should have been larger.

The variance component for lactation stages by age was also very near zero for each constituent. It was thought that, since cows in their first lactation are more persistent in milk yield than older cows, perhaps there would be an interaction between stage of lactation and age for milk composition. However, even if this is so, it apparently was an unimportant source of variation in these data. Waite et al. (1956) found the lactation curve for lactose to be flatter for cows in the first lactation than for older cows. Bartlett (1934) postulated different lactation curves for different ages.

The variance component between cows, $V(C)$, measures the totality of effects which make the average of one cow's tests differ more from the mean of all cows for that herd than would be expected from the effects of the other factors in the model on those tests. The differences between cows within herds result from both environmental and genetic causes. The environmental part of these differences would include permanent differences from one lactation to another where a cow had only one or part of one lactation and an average of these permanent lactation effects when a cow had parts of two lac-
tations. There also would be a contribution to each cow's average from a permanent environmental effect for that cow's life time. The effects of genetic differences between cows within herds would be those effects caused by the deviation of each cow's genetic constitution from the average of all the cows in that herd, i.e., these differences would be caused by additive, dominance, and epistatic deviations. The magnitude of the component for cows was surprisingly similar for all the constituents. It ranged from about 13 per cent for fat to 16 per cent for solids-not-fat. The separation of this component into environmental and genetic parts will be the basis for the genetic analysis.

The error or, perhaps more correctly, the residual component, V(E), is a conglomerate of what remained after the variance for all the other sources was removed. It contains most of the interactions which were not included in the model, although parts of some of these already may have been removed with the other sources of variation. To the extent that these ignored sources of variation are not distributed at random over the sources included in the model the estimated components are biased. In addition to the ignored sources of variation, the error component contains variance between tests by the same cow caused by unknown factors and some variance from the chemical determinations not being completely repeatable on the same milk sample. The relative magnitudes of the
different components for error ranged from about 7 per cent for the total solids to 17 per cent for solids-not-fat. One reason \( V(E) \) was larger for solids-not-fat than for the other constituents is that the contribution from "determination error" is the sum of that for fat and total solids.

Reduced Analysis of Variance and Analysis of Covariance

To complete the genetic analysis, some method of adjusting the individual tests was needed. The presence of the large interaction of months with herds for protein and solids-not-fat complicated the estimation of these adjustment factors. It was decided to separate the herd-month interaction into two parts; one part for breeds by months and one for herds by months within breeds. If most of this interaction were in the breed-month part, the adjustment factors would be much easier to estimate. It was also of interest to separate the total variance between herds into variance between herds within breeds and variance between breeds.

The several components of variance in Table 6 which had relative magnitudes of less than 2.5 per cent for every constituent were not considered in this analysis. Because it simplified the computations considerably, if the size of the component for at least one constituent was above 2.5 per cent, all of the constituents for that source were included. The
figure of 2.5 per cent is, of course, arbitrary. However, the model had to be reduced and for these data this seemed the best figure to use, for the separation. To further illuminate the relationship between milk composition and factors affecting it, the components of covariance and correlations were also calculated. The sum of the covariance components is defined the same as the sum of the variance components, i.e. it is the total covariance expected in a population where every composite sample came from a different breed, herd, month, etc. The covariance components, like the variance components, were calculated by equating the mean cross-products to their expectations. The correlations were derived by dividing the covariance component between two constituents by the geometric mean of the two appropriate variances for that source. The correlations measure the extent to which the effects of a particular source for different milk constituents vary together linearly. The components of variance and covariance are presented in Table 7, and their relative magnitudes and correlations in Table 8. The transformed data were not included in this analysis.

The effects are as defined before, and the same discussion is pertinent for the assumptions. The model for this analysis was:

\[ x_{ijklpqrs} = \mu + b_i + m_j + l_k + a_p + h_{iq} + (bm)_{ij} + (mh)_{ijq} + c_{iqr} + e_{ijklpqrs} \]
### 7. Components of variance and covariance for reduced analysis

<table>
<thead>
<tr>
<th>s of fent</th>
<th>Components of variance</th>
<th>Components of covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D.f.</td>
<td>Fat</td>
</tr>
<tr>
<td>5, s with-</td>
<td>4</td>
<td>.915</td>
</tr>
<tr>
<td>reeds</td>
<td>18</td>
<td>.066</td>
</tr>
<tr>
<td>s, siation</td>
<td>11</td>
<td>.059</td>
</tr>
<tr>
<td>V(A) s x h</td>
<td>9</td>
<td>.062</td>
</tr>
<tr>
<td>s x h, s with-</td>
<td>39</td>
<td>-.002</td>
</tr>
<tr>
<td>reeds) nths,</td>
<td>132</td>
<td>.043</td>
</tr>
<tr>
<td>s with-</td>
<td>793</td>
<td>.175</td>
</tr>
<tr>
<td>erds,</td>
<td>3416</td>
<td>.174</td>
</tr>
<tr>
<td>Source of component</td>
<td>Components of variance as percentage of &quot;total variance&quot;</td>
<td>Correlations between sources</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td></td>
<td>D.f.</td>
<td>Fat</td>
</tr>
<tr>
<td>Breeds, V(B)</td>
<td>4</td>
<td>61.7</td>
</tr>
<tr>
<td>Herds within breeds, V(H)</td>
<td>18</td>
<td>.43</td>
</tr>
<tr>
<td>Months, V(M)</td>
<td>11</td>
<td>3.9</td>
</tr>
<tr>
<td>Lactation stages, V(L)</td>
<td>9</td>
<td>4.0</td>
</tr>
<tr>
<td>Ages, V(A)</td>
<td>9</td>
<td>.5</td>
</tr>
<tr>
<td>Breeds x months, V(HM)</td>
<td>39</td>
<td>-</td>
</tr>
<tr>
<td>(Herds within Breeds) x months, V(HM)</td>
<td>132</td>
<td>2.8</td>
</tr>
<tr>
<td>Cows within herds, V(C)</td>
<td>793</td>
<td>11.4</td>
</tr>
<tr>
<td>Error, V(E)</td>
<td>3446</td>
<td>11.4</td>
</tr>
<tr>
<td>Source of component</td>
<td>D.f.</td>
<td>Components of variance and covariance</td>
</tr>
<tr>
<td>---------------------</td>
<td>------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lactose- FQ</td>
</tr>
<tr>
<td>Breeds, V(B)</td>
<td>4</td>
<td>.0061</td>
</tr>
<tr>
<td>Herds within breeds, V(H)</td>
<td>18</td>
<td>.0081</td>
</tr>
<tr>
<td>Months, V(M)</td>
<td>11</td>
<td>.0036</td>
</tr>
<tr>
<td>Lactation stages, V(L)</td>
<td>9</td>
<td>.0002</td>
</tr>
<tr>
<td>Ages, V(A)</td>
<td>9</td>
<td>.0052</td>
</tr>
<tr>
<td>Breeds x months, V(BM)</td>
<td>39</td>
<td>-</td>
</tr>
<tr>
<td>(Herds within breeds) x months, V(HM)</td>
<td>132</td>
<td>.0432</td>
</tr>
<tr>
<td>Cows within herds, V(C)</td>
<td>793</td>
<td>.0257</td>
</tr>
<tr>
<td>Error, V(E)</td>
<td>31416</td>
<td>.0794</td>
</tr>
</tbody>
</table>
The interpretations of the different components are the same as before, except that the effect of breeds is removed from differences between herds, and the effect of the breed-month interaction is removed from the herd-month interaction. Table 7 shows the smallness of the breed-month interaction components. Since the breed-month component was very near zero for each constituent, this source definitely was not the cause of the considerable herd-month components in Table 5.

The divarication of the differences between herds into differences between breeds and between herds within breeds showed that the large components between herds appearing in Table 5 were caused mostly by breed differences. The main reason the components for breeds and herds within breeds in Table 7 do not sum to the appropriate component between herds in Table 5 is because the precision of estimation is increased. This is analogous to the separation of a mean square into individual degrees of freedom or orthogonal comparisons. The other variance components have approximately the same magnitudes in Tables 5 and 7.

It perhaps is of value to illustrate some simple relationships among the variance and covariance components of Table 7. Since solids-not-fat were calculated by difference it is an identity that, \( F + S = T \). Therefore, some of the components in Table 7 need not have been calculated directly.
The unnecessary calculations were performed as a check. For instance, \( V(T) = V(F + S) = V(F) + V(S) + 2\text{Cov}(FS) \),

\[
\text{Cov}(FS) = \text{Cov}(F-T) = \text{Cov}(FT) - V(F) \text{ etc.}
\]
The small discrepancies from these relations in Table 7 are due to rounding errors. For example, the covariance component between fat and solids-not-fat for cows within herds (.0600) is approximately .236-.175.

Two parts, lactose and ash, of the total solids of milk are missing from this analysis. To aid in the interpretation of Table 8 it was decided to derive the corresponding components for lactose plus ash. Since ash is relatively constant compared to the other milk constituents, the lactose-ash portion will behave primarily as lactose. It should be realized that lactose-ash is estimated by a difference involving three separate determinations. However, the determination errors probably are random in nature and if they are random, the estimate of lactose-ash is unbiased. The main effect of random determination errors would be to make the error component of variance a bit larger than it would have been if the lactose-ash were estimated more accurately. The components of variance and covariance and their corresponding percentages and correlations for lactose-ash \((Q)\) are given in Table 9. The following relations were used to compute Table 9. Since \( Q = S - P \),
\[ V(Q) = V(S-P) = V(S) + V(P) - 2 \text{Cov}(SP) \]

\[ \text{Cov}(FQ) = \text{Cov}(F(S-P)) = \text{Cov}(FS) - \text{Cov}(FP) \]

\[ \text{Cov}(PQ) = \text{Cov}(P(S-P)) = \text{Cov}(PS) - V(P) \]

\[ \text{Cov}(SQ) = \text{Cov}(S(S-P)) = V(S) - \text{Cov}(SP) \]

\[ \text{Cov}(TQ) = \text{Cov}(T(S-P)) = \text{Cov}(ST) - \text{Cov}(PT) \]

Table 9 shows that the lactose-ash portion in some respects behaves quite differently from the other constituents. Only 3.6 per cent of the total variance of lactose-ash was in the component for breeds, whereas this source contained from 50 to 65 per cent of the total for the other constituents. Other large differences were the approximately zero component for lactation stages, the 25.2 per cent value for interaction of herds within breeds by months, and the comparatively larger percentage (46.3 per cent) in the error component.

Several correlations in Table 8 and 9 are larger than unity. Where this discrepancy is really large, at least one of the variance components involved in the correlation was practically zero. If one of the variance components actually were zero, the corresponding correlation would be indeterminate.

The large correlations between the milk constituents for breeds merely illustrate the well-known fact that if the
breeds are ordered by fat content they will also be in order, or nearly so, for the other constituents. Apparently the lactose-ash portion varies more independently of the other constituents between herds within breeds than it does between breeds. This is reflected in the smaller correlations between herds within breeds of fat, protein, and total solids with lactose-ash and, to a lesser extent, with solids-not-fat. This suggests strongly that lactose-ash is affected more by managemental conditions which vary from herd to herd than are the other constituents.

The negative correlations between lactose-ash and the other constituents for stages of lactation have as one factor in their denominator the square root of the extremely small variance component for lactose-ash and are, therefore, subject to large sampling errors. However Waite et al. (1956), Politiek (1956), and others have found a negative relationship between lactose and the other constituents for stages of lactation. All the other correlations found here for stages of lactation, months, and ages are very near unity.

The herd within breed by month interactions supposedly are caused almost wholly by environmental factors such as ration, temperature, and managemental fluctuations which are different from herd to herd. Since the herds usually were sampled only once a month, these factors are peculiar to one
day in that month. The large negative correlation of -.68 in Table 9 between protein and lactose-ash shows clearly that the herd within breed by month interaction deviations for these two constituents tend to vary inversely. This negative correlation results in the same interaction component of variance being relatively less important for solids-not-fat than for either protein or lactose-ash. The interaction effects for fat content appear to be nearly independent of those for solids-not-fat, protein and lactose-ash as indicated by the correlations of .13 and -.03 of fat with protein and lactose-ash.

Again, the correlations between cows among the constituents are large except those for lactose-ash with protein or fat. The correlations for the error source are all lowered somewhat by the automatic negativity of the correlations between component parts of a constant. For example, if total solids percentage were correlated with water percentage in individual samples the result would be minus unity. The negative relations among the component parts of total solids are not expected to be nearly so extreme, however, because it is the total milk and not total solids which is held constant. The only consequence of this tendency towards negativity among the error correlations is that one must be careful not to draw sweeping biological conclusions from relations which may be forced on the data by the mathematics involved. Here, the
error correlations among fat, lactose-ash and protein are of the same general magnitudes as the correlations among the same constituents for herd within breed by month interaction. The latter correlations would not be affected much, if at all, by the automatic negativity. Since the correlations for both of these sources are mostly caused by environmental effects, the automatic negativity probably is not a large factor in making the error correlations more negative. Intra-cow correlations found between fat and solids-not-fat were -.02 by Harvey et al. (1954); and -.14 by Musgrave et al. (1958). Tocher (1927) found a negative relation between fat and solids-not-fat for day-to-day variations of herd milks from the same herd.

**Relationship between Solids-Not-Fat and Fat**

Herrmann (1954) in a literature review summarized some 70 different linear equations for predicting solids-not-fat from either fat alone or fat plus specific gravity. Several milk pricing formulas are in use which assume some definite relation between non-fat solids and fat. The widely used formulae for fat-corrected milk also assume this relation to be constant. Especially in recent years, more recognition has been given to the fact that the relation found between solids-not-fat and fat may differ according to the structure of the population sampled. Consequently, it seemed of value to demonstrate how the variance and covariance components in
Table 7 combine in the sum of their effects on these two constituents.

If the percentages given in Tables 8 and 9 are converted to decimal fractions, and if the model was correct in assuming that the different sources were uncorrelated, they become squared path coefficients, Wright (1921). For an excellent discussion of the relation between path coefficients and the analysis of variance, see Hazel et al. (1943). Using the relations hypothesized in the path coefficient diagram in Figure 6, the total correlation between fat and solids-not-fat can be derived for the population defined as it was for the sum of the variance and covariance components.

The derived total correlation, \( r_{FS} \), then is

\[
r_{FS} = b_F r_{BFBS} b_S + h_F r_{HPHS} h_S + \ldots + e_F r_{BEBS} e_S,
\]

where \( b_F \) is the square root of the fraction of the total variance in fat due to breeds, \( r_{BFBS} \) is the correlation between fat and solids-not-fat for the source of breeds, \( b_S \) is the square root of the fraction of the total variance in solids-not-fat for the source of breeds and etc. The derived total correlation, \( r_{FS} = .695 = (\sqrt{.617}) (.93) (\sqrt{.502}) + (\sqrt{.043}) (.66) (\sqrt{.034}) + \ldots + (\sqrt{.114}) (-.12) (\sqrt{.166}) \). The correlation calculated from the original data was .660. The small discrepancy between the derived and original correlations results
Figure 6. Path coefficient diagram hypothesizing the complete determination of fat and solids-not-fat contents by the effects of breeds, $B_F$ and $B_S$, herds within breeds, $H_F$ and $H_S$, and etc. The correlations between the two constituents for the different sources are $r_{B_FB_S}$, etc. and the path coefficients from the different sources to the constituents are $b_F$, $b_S$ . . . eg.
from the model not being exactly correct in assuming the different sources to be independent, and/or from considering each individual milk sample as coming from a different breed, herd, month, etc. Whereas, in fact, many of these samples came from the same source.

Corresponding to the correlation between fat and solids-not-fat for each source, there exists a regression of solids-not-fat on fat for each source computed by dividing the covariance between fat and solids-not-fat by the variance of fat for each particular source. The total regression of solids-not-fat can be derived in a manner similar to the derivation of the total correlation. To derive the total regression, the regression of solids-not-fat on fat for each source is multiplied by the squared path coefficient for that source and summed over all sources.

Apparently the interactions, other than those for herds within breeds by months, were relatively small (Table 6). Thus, the variance and covariance components in Table 7 are expected to be approximately the same as would be found for a population of samples from herds all of the same breed or even for a population from a single herd. Except, of course, in these latter situations some of the variance and covariance components would be missing. Table 10 shows the correlations
and regressions of solids-not-fat on fat for each source and the corresponding functions of path coefficients which are the appropriate weights for three situations; (a) the data as they were collected, (b) on an intra-breed basis and (c) on an intra-herd basis.

The total correlation weights, $w_j$, in Table 10 are the products of the two appropriate path coefficients. Since the total variances were the only statistics expected to change for the intra-breed weights, $w_k$ was derived by $w_j \left( \sqrt{\frac{1}{1-.617}} \right) \left( \sqrt{\frac{1}{1-.502}} \right)$; where .617 and .502 are the expected relative decreases in the total variance when differences between breeds are eliminated. The $z_k$ values were computed by $z_j \left( \frac{1}{1-.617} \right)$. The $w_1$ and $z_1$ quantities were derived similarly.

It is immediately apparent from Table 10 that no constant relationship exists between solids-not-fat and fat. Table 10 also indicates some of the reasons for the considerable disagreement over the magnitude of this relationship. The probable reason for the discrepancies between the derived correlations and regressions and those actually calculated has been discussed previously. The pertinent information in Table 10 is that the causes for the differences between the derived regressions are the same as those which caused the regressions calculated in the usual manner to differ. The intra-breed and intra-herd regressions are nearly alike because elimination of the large regression between herds and the small re-
gression for herd-month interaction approximately balanced in the net effect on the intra-herd regression.

The regressions discussed previously were for populations of single samples from individual cows. Many regressions of solids-not-fat on fat have been calculated from bulked or herd milks. It perhaps is of interest to derive from the present data an approximate corresponding regression for herd milks. To do this, the assumption is necessary that there would be no appreciable difference between the variance of actual herd milks and the variance of estimated herd milks where the cows' tests are weighted equally.

With 20 cows per herd the regressions in Table 10 would not be expected to change, but the weights for the total regression would change because the variance components for fat, \( V(L), V(A), V(C) \) and \( V(E) \), in Table 7 would be reduced to approximately one twentieth of their original values (still assuming random variables). The derived regression of solids-not-fat on fat for herd milks from herds with 20 cows each was 0.415. The calculated regression from herd-month subclass averages of the original data was 0.408.

The object here was not to estimate the regression of non-fat solids on fat, but rather to clarify the reasons that the regression for herd milks has usually been found to be of
<table>
<thead>
<tr>
<th>Source of variance or covariance</th>
<th>Correlation weights, ( w_j )</th>
<th>Total correlation</th>
<th>Intra-breed correlation</th>
<th>Intra-herd correlation</th>
<th>Regression of solids on fat</th>
<th>Total regression</th>
<th>Intra-breed regression</th>
<th>Intra-herd regression</th>
<th>Herd milk regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breeds</td>
<td>( r_i )</td>
<td>( r_{ij} )</td>
<td>( z_j )</td>
<td>( z_k )</td>
<td>( z_1 )</td>
<td>( z_{ij} )</td>
<td>( z_{ik} )</td>
<td>( z_{ik} )</td>
<td>( z_{ik} )</td>
</tr>
<tr>
<td>Breeds</td>
<td>.93</td>
<td>5566</td>
<td>-</td>
<td>-</td>
<td>44</td>
<td>617</td>
<td>-</td>
<td>-</td>
<td>833</td>
</tr>
<tr>
<td>Herds within breeds</td>
<td>.66</td>
<td>.0381</td>
<td>.0872</td>
<td>-</td>
<td>.31</td>
<td>.043</td>
<td>.1123</td>
<td>-</td>
<td>.058</td>
</tr>
<tr>
<td>Months</td>
<td>1.07</td>
<td>.0286</td>
<td>.0655</td>
<td>.0801</td>
<td>.42</td>
<td>.039</td>
<td>1.018</td>
<td>.1250</td>
<td>.052</td>
</tr>
<tr>
<td>Lactation stages</td>
<td>.99</td>
<td>.0414</td>
<td>.0948</td>
<td>.1159</td>
<td>.53</td>
<td>.040</td>
<td>.1044</td>
<td>.1282</td>
<td>.003</td>
</tr>
<tr>
<td>Ages (Herds within breeds)</td>
<td>1.21</td>
<td>.0121</td>
<td>.0277</td>
<td>.0339</td>
<td>1.50</td>
<td>.005</td>
<td>.0131</td>
<td>.0160</td>
<td>.000</td>
</tr>
<tr>
<td>Cows within herds</td>
<td>.17</td>
<td>.0391</td>
<td>.0895</td>
<td>-</td>
<td>.13</td>
<td>.028</td>
<td>.0731</td>
<td>-</td>
<td>.038</td>
</tr>
<tr>
<td>Ages (Herds within breeds)</td>
<td>( x )</td>
<td>.01379</td>
<td>.3158</td>
<td>.3860</td>
<td>-.07</td>
<td>.114</td>
<td>.2977</td>
<td>.3654</td>
<td>.008</td>
</tr>
<tr>
<td>Derived coefficients</td>
<td>Total: ( \Sigma w_j r_{ij} = 695 )</td>
<td>( \Sigma w_r_{ij} = .406 )</td>
<td>( \Sigma w_{ij} = .407 )</td>
<td>( \Sigma z_{ij} = .364 )</td>
<td>( \Sigma k_{ib} = 243 )</td>
<td>( \Sigma k_{ib} = .145 )</td>
<td>( \Sigma k_{ib} = .216 )</td>
<td>( \Sigma k_{ib} = .145 )</td>
<td>( \Sigma k_{ib} = .216 )</td>
</tr>
<tr>
<td>Coefficients calculated directly</td>
<td>.660</td>
<td>.376</td>
<td>.365</td>
<td>.347</td>
<td>.225</td>
<td>.216</td>
<td>.148</td>
<td>.216</td>
<td>.148</td>
</tr>
</tbody>
</table>
the order of .4, whereas the intra-breed or intra-herd regression for individual cow's samples is apparently nearer .2 and the intra-cow regression probably is slightly negative. Jenkins and Provan (1956) also gave regressions of solids-not-fat on fat for different methods of sampling; however, their sampling methods are not described well enough to compare their regressions with these. That the correlations and regressions calculated directly and those derived by the path coefficient method agreed as well as they did indicates that the model and methods used for the analyses of variance were reasonably accurate.

Estimates of Environmental and Physiological Effects

As mentioned previously, estimates of those environmental and physiological effects which appeared important in the variance analysis were needed not only for their own value but also to adjust the data for the genetic analysis. Because of the disproportionate subclass numbers and multiple classification it was necessary to derive these estimates by the method of least squares. Hazel (1946), Henderson (1948) and others, using examples from animal breeding experiments, have described adequately the method of "fitting constants" by least squares.

A correct method for deriving the estimates to obtain unbiased estimates of the genetic parameters from the adjusted
data was given by Henderson (1953); his "Method II". To use this method the adjustment factors must be estimated jointly with the effects to be considered in the genetic analysis. This joint consideration is most easily accomplished by including in the model used to fit the constants a term for the smallest subclass containing genetic and permanent environmental effects. In this study the smallest subclasses in the genetic analysis were to be formed from the tests for each cow during one lactation. However, it was impossible to include a term for each cow's lactation because the effects of months and stages of lactation are completely confounded within each cow's lactation and, thus, their effects are inseparable. This confounding results because the different months of each cow's lactation and the months of the year in which that cow was tested follow the same sequence. If simultaneous estimation is attempted in such a situation the equations will have no unique solution. Koch (1951), and Blackwell and Henderson (1955) have encountered similar situations with the involved classifications being years and ages. In addition to the difficulties caused by the confounding, it would not have been computationally feasible to account for lactations.

Differences between lactations also contained differences between cows because each cow was represented by one or, at most, parts of two lactations. Disregarding lactations
in the least squares model will define the lactation differences almost wholly into the error component. Statistically, this causes correlations among the errors. However, if lactations can be considered as random variables, least squares still yields unbiased estimates of the other effects in the model.

Selection apparently is the only force which would invalidate the assumption of randomness of lactations. As discussed previously, selection probably has not had a chance to cause large differences between lactations (cows) because, except for fat content, the farmers had no information on which to base selections. It was not possible to investigate the extent to which selection had been practiced because data for only about one year were available for each herd. Even if selection were important, if the cows in the different age groups are distributed at random over the other effects to be estimated, these estimates would not be biased seriously. Seemingly, no reasons exist to expect this distribution to be other than random. Table 3 shows that except for ages, the amounts of variance between cows expected to be contained in the other effects are small.

Even if lactations are random variables, not including them in the model used to estimate the adjustment factors could cause later difficulties in the genetic analysis. This
is because each adjustment factor will contain an average contribution of the lactations in that particular classification. For example, consider a one way classification for ages. The model, disregarding lactations, would be

$$\mu + a_i + e_{ijk}.$$ 

However, each age adjustment factor, the mean for each age, would contain the true age effect, $\alpha_i$, plus some average contribution from lactations and errors, i.e.

$$a_i = \alpha_i + \frac{\sum l_{ij}}{n_i} + \frac{\sum e_{ijk}}{n_i}.$$ 

Assuming $l_{ij}$ and $e_{ijk}$ to be random variables, the expectations of $\sum l_{ij}/n_i$ and $\sum e_{ijk}/n_i$ are zero. However, these quantities, in fact, generally will not be exactly zero. If the data to be used in the genetic analysis are adjusted by subtracting the estimated age correction factors, the obvious consequences are that part of the error and part of the lactation effects will be removed. Henderson's "Method II" ordinarily would be used to correct the expected mean squares in the genetic analysis for the fact that part of the error was removed, but this method could not be applied to these data.

Fortunately, the seriousness of disregarding lactations is inversely proportional to the numbers of observations available to estimate the adjustment factors. The herd-month sub-
classes contained the smallest number of observations in these data. Since the average number in these subclasses was moderately large (about 20), it was concluded that disregarding lactations in the present analysis probably would not affect seriously the results obtained from the adjusted data.

The presence of the herd-month interaction caused computing problems. It appeared that the number of equations to be solved simultaneously could not be reduced extensively by absorption techniques. At this point in the analysis rough plots of the data were made to observe any tendencies of the effects to exhibit trends.

The age effects seemed generally to decrease steadily as the age increased. Biologically, age effects would be expected to follow some steady pattern. At least, irregular fluctuations would be inconsistent with any known physiological effect of aging. Solids-not-fat seemed to decrease linearly as age increased, but the other constituents appeared to decrease curvilinearly. In order to use the same model for all constituents, it was decided to fit quadratic regressions to the effects of age.

For their analysis of genetic effects, Robertson et al. (1956) discarded the tests during the first month and the last two months of the lactation. They justified this from the
variance of these tests being larger than the variance of the constituents during the middle months of lactation. Here, it appeared that if the data for the remaining 20 days of the first month of lactation were eliminated, linear regressions would fit the other nine months reasonably well. Thus, partly because the variance of the tests is larger during the first month of lactation but mostly to reduce the computations to manageable proportions, the observations for the first month of lactation were discarded and a linear regression term was placed in the model for stage of lactation.

Some of the herds had two different test dates in the same month. These test dates were separated for this analysis and each one was considered as having its own peculiar effect. This increased the number of herd-month subclasses from 203 to 217. It did not seem practical to estimate separately constants for herds, months, and herd-month interaction, so these effects were considered jointly by fitting a constant for each herd-month subclass. The absence of large herd-age and herd-lactation stage interactions in the variance analysis was the justification for fitting common regressions to all herds. After the tests for the first month of lactation were eliminated, 4,106 observations on each constituent remained for this analysis.

The model was written as,
\[ X_{ijkl} = \mu + (hm)_i + b_1s_j + b_2a_k + b_3a_k^2 + e_{ijkl} \]

where: \( X_{ijkl} \) is the content of a milk constituent in a daily composite from a cow tested at the \( k \)th age during the \( j \)th stage of lactation on the \( i \)th herd-month test date.

The transformed data were also included in this analysis. However, after the data were adjusted and the genetic analysis finished, the difference was never larger than .03 between the estimates of the genetic parameters for the transformed and original data. These differences did not appear to make the estimates for the transformed data consistently either higher or lower than those for the original data. Therefore, the transformed data will not be considered further.

The normal equations were solved by absorbing the 217 \( \mu + (hm)_i \) equations into the regression equations. The three regression equations were solved simultaneously. The constants \( b_1, b_2, \) and \( b_3 \) were then substituted in the original \( \mu + (hm)_i \) equations to obtain the \( \mu + (hm)_i \) constants. The \( (hm)_i \) effects were then separated from \( u \) by imposing the restriction, \( \sum_i (hm)_i = 0. \)

The partial regression coefficients for each constituent are presented in Table 11. These regressions measure the average change in a particular constituent for each increase of one month in stage of lactation, one year in age, or one year
Table 11. Regressions of milk constituents on age and stage of lactation

<table>
<thead>
<tr>
<th>Regression of milk constituent on</th>
<th>Fat</th>
<th>Protein</th>
<th>Solids-not-fat</th>
<th>Total solids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactation stage, $b_1$</td>
<td>0.0806 ± 0.0042</td>
<td>0.0528 ± 0.0018</td>
<td>0.0476 ± 0.0025</td>
<td>0.1282 ± 0.0055</td>
</tr>
<tr>
<td>Age, $b_2$</td>
<td>0.0177 ± 0.0184</td>
<td>0.0123 ± 0.0079</td>
<td>-0.0258 ± 0.0012</td>
<td>-0.0081 ± 0.0241</td>
</tr>
<tr>
<td>Age squared, $b_3$</td>
<td>-0.0033 ± 0.0017</td>
<td>-0.0021 ± 0.0007</td>
<td>-0.0009 ± 0.0010</td>
<td>-0.0042 ± 0.0022</td>
</tr>
</tbody>
</table>
Table 12. Distribution of tests for ages and stages of lactation

<table>
<thead>
<tr>
<th>Year of age or month of lactation</th>
<th>Frequencies of tests</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>Lactation stage</td>
</tr>
<tr>
<td>1</td>
<td>184</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>924</td>
<td>528</td>
</tr>
<tr>
<td>3</td>
<td>910</td>
<td>529</td>
</tr>
<tr>
<td>4</td>
<td>588</td>
<td>507</td>
</tr>
<tr>
<td>5</td>
<td>463</td>
<td>496</td>
</tr>
<tr>
<td>6</td>
<td>364</td>
<td>473</td>
</tr>
<tr>
<td>7</td>
<td>252</td>
<td>454</td>
</tr>
<tr>
<td>8</td>
<td>146</td>
<td>438</td>
</tr>
<tr>
<td>9</td>
<td>126</td>
<td>365</td>
</tr>
<tr>
<td>10</td>
<td>149</td>
<td>316</td>
</tr>
</tbody>
</table>

in age squared. Table 12 gives the number of samples in each age and stage of lactation group.

Figures 7 to 11 show the regressions in Table 11 for lactation stage and the means from the adjusted data for each month of lactation. The means were derived from the data after each observation was adjusted by subtracting the appropriate constants for \((hn)_{i}, b_{2a_{k}}, \text{ and } b_{3a_{k}^2}\). The adjusted means for each month of lactation can not be regarded as the exact constants which would have been obtained if the
effect of each month of lactation had been considered ab initio. However, these adjusted means can be considered as the constants which would be obtained from an iterative solution of the least squares equations where the expected values from the regressions $b_1$, $b_2$, and $b_3$ were substituted for these effects during the previous round of iteration. Except perhaps for lactose-ash in late months of lactation, the regressions do fit the data reasonably well. Consequently, for all practical purposes, the adjusted means are equivalent to exact least squares estimates. The scales used for the "dependent" variables in Figures 7 to 11 are roughly equal to the standard deviation for each constituent.

It is apparent from Figures 7 to 11 that the effect of stage of lactation on fat and total solids was essentially linear in these data. However, the protein content seemed to increase a bit more rapidly from the second to fifth month of lactation than in later months. This pattern agrees with the results of Politiek (1956); however, Waite et al. (1956) and others stated that both crude protein and casein increase most rapidly during the latter months of lactation.

The curve for the influence of lactation stage on solids-not-fat agrees in general with most of the reports. Apparently the solids-not-fat content reaches a low point in the second month of lactation, increases steadily until about the fifth month, levels off slightly until the seventh month and then
increases steadily again from the eighth to tenth month. Bailey (1952c) demonstrated that the increase during the last three months did not occur in non-pregnant cows. However, as Waite et al. (1956) indicated, the great majority of cows are pregnant during this time.

The influence of lactation stage on lactose-ash shown in Figure 11 does not agree with the hypothesis that lactose-ash behaves essentially as lactose. Lactose content commonly has been reported as decreasing steadily over the lactation period. The ash content may have caused the increase during the latter part of the lactation appearing in Figure 11. Jacobson and Wallis (1938) noted an increase of approximately .08 per cent in ash during the last three months of lactation.

Figures 12 to 16 show the regressions and adjusted means for ages. The adjusted means were derived in the same manner as those for lactation stages. The general pattern of the relation between age and milk composition agrees with most of the other reports. A striking feature of the age graphs is the large positive deviations from the regressions for nine year old cows. A biological explanation for this is not obvious. Apparently these deviations resulted from chance in sampling, although the probability of chance as a cause seems small with 126 observations at this age. That this large deviation occurred for all the constituents was to be expected
Figure 7. Relation between stage of lactation and fat content

Figure 8. Relation between stage of lactation and total solids content
Figure 9. Relation between stage of lactation and protein content

Figure 10. Relation between stage of lactation and solids-not-fat content
Figure 11. Relation between stage of lactation and lactose-ash content

Figure 12. Relation between age and lactose-ash content
Figure 13. Relation between age and protein content

Figure 14. Relation between age and solids-not-fat content
Figure 15. Relation between age and fat content

Figure 16. Relation between age and total solids content
from the correlations of nearly unity obtained between the age deviations of the constituents in the covariance analysis.

Fat content commonly has been reported to decrease linearly as age increases, Gowen (1924). However some have found higher values for second than for first lactations; as for example, Johansson and Hansson (1940). The standard errors in Table 11 show that the quadratic regression on age is significant for protein content but not for non-fat solids, and is on the border line of being significant for fat and total solids contents. The linear regression for solids-not-fat is \(-.040 \pm .002\) per cent for each increase in one year in age. The hypothesis that most of the decrease in solids-not-fat content due to age is caused by decreased lactose is supported by these data.

Since the effects of herds, months and herd-month interaction were estimated jointly, exact least-squares estimates of these effects were not available. However approximate estimates can be obtained from the \(\mu + (hm)_i\) constants. Strictly speaking, unbiased estimates could be derived in this manner only if all the herd-month subclasses were filled. Nevertheless, this seemed to be the best practical procedure even if it did yield slightly biased estimates.

To estimate the herd means, the \(\mu + (hm)_i\) constants were averaged for each herd. The herd means were adjusted to cor-
respond roughly to the means expected for lactation averages. This was accomplished by adjusting to a basis of cows averaging four years in age and producing during the fifth month of lactation. The estimated herd means are given in Table 13. The estimated breed means, Table 14, are weighted averages; the weights being the number of tests for each herd.

Of the 23 herds, four of the larger ones had been sampled each month for 12 consecutive months. These four herds were divided between the two areas which differed most in climate and, perhaps, in feeding conditions. It was decided to illustrate the herd-month subclass means for protein and solids-not-fat with the constants for these herds. The constants, expressed as deviations from the pertinent herd means, are given in Figure 17.

It is clear from Figure 17 that no sharp increase in protein content, such as noted by Waite et al. (1956) and Politiek (1956), occurred soon after the cows were first put to pasture in the spring. This increase would have been expected probably in April for area I and in June for area III. The difference between winter and summer months for non-fat solids, and to a lesser extent for protein, was larger in area III than area I. This larger decrease could have been due to the warmer summer and/or colder winter in area III; however, it also could have been due to differences between the two areas in other environ-
Table 13. Estimated herd means, adjusted to approximate lactation averages

<table>
<thead>
<tr>
<th>Herd</th>
<th>Number of tests</th>
<th>Fat</th>
<th>Protein</th>
<th>Solids-not-fat</th>
<th>Total solids</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 1</td>
<td>153</td>
<td>3.99</td>
<td>3.22</td>
<td>8.77</td>
<td>12.76</td>
</tr>
<tr>
<td>B 1</td>
<td>210</td>
<td>4.02</td>
<td>3.61</td>
<td>9.34</td>
<td>13.36</td>
</tr>
<tr>
<td>B 2</td>
<td>254</td>
<td>3.76</td>
<td>3.50</td>
<td>9.27</td>
<td>13.03</td>
</tr>
<tr>
<td>G 1</td>
<td>208</td>
<td>4.87</td>
<td>3.57</td>
<td>9.41</td>
<td>14.28</td>
</tr>
<tr>
<td>G 3</td>
<td>149</td>
<td>4.66</td>
<td>3.64</td>
<td>9.40</td>
<td>14.06</td>
</tr>
<tr>
<td>G 4</td>
<td>214</td>
<td>4.96</td>
<td>3.60</td>
<td>9.26</td>
<td>14.22</td>
</tr>
<tr>
<td>G 5</td>
<td>215</td>
<td>5.12</td>
<td>3.60</td>
<td>9.42</td>
<td>14.54</td>
</tr>
<tr>
<td>H 1</td>
<td>127</td>
<td>3.45</td>
<td>2.94</td>
<td>8.38</td>
<td>11.83</td>
</tr>
<tr>
<td>H 2</td>
<td>318</td>
<td>3.64</td>
<td>3.06</td>
<td>8.78</td>
<td>12.42</td>
</tr>
<tr>
<td>H 3</td>
<td>245</td>
<td>3.51</td>
<td>3.03</td>
<td>8.64</td>
<td>12.15</td>
</tr>
<tr>
<td>H 4</td>
<td>190</td>
<td>3.51</td>
<td>3.24</td>
<td>8.63</td>
<td>12.14</td>
</tr>
<tr>
<td>H 5</td>
<td>127</td>
<td>3.22</td>
<td>2.99</td>
<td>8.43</td>
<td>11.65</td>
</tr>
<tr>
<td>H 6</td>
<td>199</td>
<td>3.54</td>
<td>3.04</td>
<td>8.59</td>
<td>12.13</td>
</tr>
<tr>
<td>J 1</td>
<td>216</td>
<td>6.26</td>
<td>4.04</td>
<td>9.76</td>
<td>16.02</td>
</tr>
<tr>
<td>J 2</td>
<td>161</td>
<td>5.19</td>
<td>3.82</td>
<td>9.32</td>
<td>14.51</td>
</tr>
<tr>
<td>J 3</td>
<td>108</td>
<td>5.82</td>
<td>3.91</td>
<td>9.67</td>
<td>15.49</td>
</tr>
<tr>
<td>J 4</td>
<td>90</td>
<td>5.44</td>
<td>3.93</td>
<td>9.60</td>
<td>15.04</td>
</tr>
<tr>
<td>J 5</td>
<td>287</td>
<td>5.30</td>
<td>3.88</td>
<td>9.47</td>
<td>14.77</td>
</tr>
<tr>
<td>J 6</td>
<td>124</td>
<td>5.64</td>
<td>3.88</td>
<td>9.77</td>
<td>15.41</td>
</tr>
<tr>
<td>J 7</td>
<td>120</td>
<td>5.35</td>
<td>3.88</td>
<td>9.43</td>
<td>14.78</td>
</tr>
<tr>
<td>J 8</td>
<td>151</td>
<td>5.61</td>
<td>3.89</td>
<td>9.67</td>
<td>15.28</td>
</tr>
<tr>
<td>J 9</td>
<td>104</td>
<td>4.95</td>
<td>3.82</td>
<td>9.47</td>
<td>14.42</td>
</tr>
</tbody>
</table>

Table 14. Estimated breed means, adjusted to approximate lactation averages

<table>
<thead>
<tr>
<th>Breed</th>
<th>Number of tests</th>
<th>Fat</th>
<th>Protein</th>
<th>Solids-not-fat</th>
<th>Total solids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayrshire</td>
<td>153</td>
<td>3.99</td>
<td>3.22</td>
<td>8.77</td>
<td>12.76</td>
</tr>
<tr>
<td>Brown Sw.</td>
<td>464</td>
<td>3.88</td>
<td>3.55</td>
<td>9.30</td>
<td>13.18</td>
</tr>
<tr>
<td>Guernsey</td>
<td>922</td>
<td>4.95</td>
<td>3.59</td>
<td>9.33</td>
<td>14.28</td>
</tr>
<tr>
<td>Holstein</td>
<td>1206</td>
<td>3.51</td>
<td>3.06</td>
<td>8.62</td>
<td>12.13</td>
</tr>
<tr>
<td>Jersey</td>
<td>1361</td>
<td>5.53</td>
<td>3.90</td>
<td>9.57</td>
<td>15.10</td>
</tr>
</tbody>
</table>
mental conditions. No way was apparent for separating the different possible causes.

The monthly averages have some interest since they indicate how the milk composition varies from month to month over the whole material. Figure 18 shows the average of the $(hm)_i$ constants for each constituent by months. Fat, total solids and solids-not-fat contents usually were higher in the fall and winter and lower during the spring and summer months. This pattern agrees generally with the results of other investigators. No general pattern seemed apparent for protein content.

Heritabilities, Repeatabilities and Genetic and Phenotypic Correlations

The basic unit of dairy cattle production records is a cow's production during one lactation. Although slight differences exist in the length of lactation used, it has become common practice to report statistics in terms of what happens with single lactations. Fortunately, this allows more accurate comparisons of the results of different workers than would be possible otherwise. Therefore, if single lactations are not used it is important, when it makes a difference, to adjust the statistics to a single record basis.

Ordinarily, the lactation record for fat percentage is the weighted average test of daily composites taken at approx-
Figure 17. Sample of herd-month constants for protein and solids-not-fat contents
Figure 18. Monthly averages of herd-month constants
imately monthly intervals during the first ten months of lactation. Regardless of whether the tests are multiplied by the number of days in the month, added over months for a 305 day record and the fat content then calculated, the basic datum is the average of ten tests. The lactation record in this study is defined as the average of nine tests because the tests for the first month of lactation were discarded. Not enough complete lactations were available in the present data to warrant estimating genetic parameters from these alone. However, assuming that the analysis of the factors affecting the individual tests separated fairly accurately the components contributing to the total variance, the necessary knowledge is available to reconstruct what would have happened if, in fact, complete lactation averages had been used.

If the data were from cows all with complete lactation averages and tested during the same months, the average effects of stage of lactation and month of year would contribute nothing to the differences among cows. To the extent that the effects of lactation stage and month of year differed from cow to cow these effects would be included in the differences among cows. Further, if the cows were of the same breed and farm, differences among them would include no differences from breeds, herds, or herd-month interaction. Thus, in using the individual tests to estimate the parameters of the
intra-herd lactation averages, the variance due to breeds, herds within breeds, stage of lactation, months and herd-month interaction needs to be eliminated.

All of a cow's tests during one lactation occur at the same year of age; therefore, the effects of differences in ages are included fully in the differences between lactation averages of cows not the same age. Since the variance among lactation averages within herds is much smaller than the variance among individual tests, differences due to ages become more important among those averages than was indicated previously. The differences due to ages then also should be removed for the genetic analysis.

To aid in presenting the estimated phenotypic and genetic parameters, some definitions will be useful. In a population of samples from the same herd and breed define:

\( V(G) \) as the variance caused by deviations among the additive portions of the genotypes of cows, i.e. variance among "additive breeding values".

\( V(D) \) as the variance due to dominance deviations among the genotypes, i.e. such variance as occurs because the phenotype of the heterozygote is not exactly half way between the phenotypes of the two homozygotes for any particular pair of alleles.

\( V(I) \) as the variance caused by epistatic deviations
among the genotypes, i.e. interaction among non-allelic genes.

\( V(Z) \) as the variance due to the environmental effects which are alike for each cow during her lifetime.

\( V(R) \) as the variance caused by the environmental effects which are alike throughout one lactation but differ from lactation to lactation for the same cow.

\( V(W) \) as the variance due to random effects which differ from test to test within a lactation of a particular cow.

\( V(C) \) as the variance among lactation averages of nine equally weighted tests per cow. If no interaction or correlation between genotype and environment exists; 

\[
V(C) = V(G) + V(D) + V(I) + V(Z) + V(R) + \frac{V(W)}{9}
\]

\( V(\bar{G}) \) as the variance among the genotypes, 

\[
V(\bar{G}) = V(G) + V(D) + V(I).
\]

\( V(E) \) as the environmental variance plus the epistatic and dominance variances, 

\[
V(E) = V(D) + V(I) + V(Z) + V(R) + \frac{V(W)}{9}
\]

Heritability \((h^2)\) can be defined as the regression of an individual's genotype on its phenotype. For so-called quantitative traits the genotype is unknown but can be predicted with a regression equation where the individual's phenotype
is used as the "independent" variable. The function of the genotype to be predicted can be taken as \( H \) or, more usefully for most purposes, as \( G \). Because of the part-whole relationship between the genotype and phenotype, heritability becomes the ratio of the pertinent genetic variance to the total or phenotypic variance.

Lush (1948) termed \( \frac{V(H)}{V(C)} \) heritability in the "broad sense" and \( \frac{V(G)}{V(C)} \) heritability in the "narrow sense". The regression of additive breeding value on phenotype (heritability in the narrow sense) is ordinarily the most useful, because in a random mating population it is proportional to the fraction of differences between the phenotypes of parents expected to be recovered in their offspring.

All methods of estimating heritability utilize the fact that relatives have genes in common and therefore resemble each other more than do unrelated members of the population. Provided that the relatives do not also resemble each other more because of a common environment, estimates of the reduction in variance caused by the common genes can be obtained by subtracting the variance between related animals from that between unrelated animals, or by calculating the covariance between relatives. In a random mating population, these estimates have expectations of the coefficient of relationship times \( V(G) \), plus part of \( V(I) \) and sometimes part
of $V(D)$. For a method of deriving the expected amounts of $V(I)$ and $V(D)$ for differently related families in random mating populations, see Kempthorne (1957).

In data from dairy cattle, the most numerous groups of relatives are daughter-dam pairs and paternal half sister families. Heritabilities were estimated from both daughter-dam pairs and paternal half sister groups in the present study.

Hazel (1943) advanced a method to separate a phenotypic correlation into two component parts, an environmental correlation ($r_{EXEY}$) and a genetic correlation ($r_{GxGy}$). Robertson et al. (1956) and Tyler (1958) discussed the difference between genetic and phenotypic correlations for the specific case of milk constituents. Nevertheless, considering some of the statements in the literature, the reasons these correlations can be different seem to deserve reiteration.

The phenotypic correlation between two traits measures the accuracy of predicting the phenotype of one trait from that of another trait in the same individual. But that which is needed to predict the response of one trait when the selections are based on the phenotypes of another trait is a function of the genetic correlation. Both the phenotypic and genetic correlations are needed for combining optimally sever-
al traits into a selection index.

Because dominance deviations and, for the most part, epistatic deviations are not transmitted to the offspring, under random mating, the correlation between the additive genetic deviations of two traits is the most useful. Letting \( x \) and \( y \) denote two different traits, the phenotypic correlation \( r_{xy} = \frac{\text{Cov}(G_x + E_x)(G_y + E_y)}{\sqrt{V(G_x + E_x)V(G_y + E_y)}} \). Assuming zero correlation between the \( G \)'s and the \( E \)'s, \( r_{xy} = \frac{\text{Cov}(G_xG_y) + \text{Cov}(E_xE_y)}{\sqrt{V(G_x)V(G_y)}} \)

\[
= \frac{\text{Cov}(G_xG_y)}{\sqrt{V(G_x)V(G_y)}} \frac{1}{\sqrt{V(G_x)V(G_y)}} + \frac{\text{Cov}(E_xE_y)}{\sqrt{V(E_x)V(E_y)}} \frac{1}{\sqrt{V(E_x)V(E_y)}} \\
= \sqrt{h_x^2 h_y^2 + r_{E_xE_y} \left( 1 - h_x^2 \right) \left( 1 - h_y^2 \right)} 
\]

Lush (1948) discussed the possible causes of genetic correlations. When a population has been mating at random for several generations and when selection has been of the same general magnitude and in the same direction for the different families within each subgroup considered, the only plausible cause of genetic correlations when the variations within subgroups are analyzed is pleiotropy. Pleiotropy is defined as the phenomenon of the same genes simultaneously affecting different characters. In the present study a subgroup is a herd, and no reasons are apparent for doubting that mating was almost random and selection pressures were
about equal for different families within herds.

Environmental correlations probably result from the same environmental incident happening to an individual and tending to happen to it all over, if it happens at all. Also, the dominance deviations and most of the epistatic deviations which may simultaneously affect the two pertinent traits are included in the "environmental" correlation as it is defined in this analysis.

Since one is able to observe only the phenotypes of two different traits, the environmental and genetic effects cannot be separated in the same animal. However, as proposed by Hazel (1943), the fact that relatives have genes in common can be utilized to estimate the genetic correlations. This method, in order to be valid, assumes that the environmental correlation is zero between trait x in one relative and trait y in the other. But, the structure of these data seem to make that assumption reasonably safe.

Repeatability is the correlation between random records on the same cow. Lush (1945) discussed extensively the nature and use of repeatability. If repeatability is high, only a few records need to be taken to estimate real producing ability accurately, and vice versa. This is because the random environmental variation is, on the average, divided by n,
the number of records per animal, in the average of \( n \) records. If the random environmental portion of the total variance is large, repeatability is low and increasing \( n \) does much to increase the correlation between the average of the observations and the real producing ability.

In this investigation, two different repeatabilities can be estimated. These are the correlation between tests taken during the same lactation of a cow and the correlation between estimated lactation averages of the same cow.

For the paternal half sib analysis, 4015 tests on each milk constituent from 793 daughters of 232 sires were available. The average effects of age and of stage of lactation were removed by adjusting the data with the regressions shown in Table 11. The effects of breeds, herds, months and herd-by-month interactions were considered by including terms for herds and for months within herds in the model used to estimate the pertinent components of variance and covariance. As in the preceding analyses, the mean squares and cross-products were equated to their expectations and the resulting equations were solved for the components. The nature of the data allowed the analysis of variance to take the familiar hierarchical arrangement. The estimated variance and covariance components are given in Tables 15 and 16.
Table 15. Components of variance from paternal half sibs

<table>
<thead>
<tr>
<th>Source of variance component</th>
<th>D.f.</th>
<th>Fat</th>
<th>Protein</th>
<th>Solids-not-fat</th>
<th>Total solids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires within herds, V(J)</td>
<td>209</td>
<td>.0260</td>
<td>.0045</td>
<td>.0075</td>
<td>.0483</td>
</tr>
<tr>
<td>Cows within sires, V(K)</td>
<td>561</td>
<td>.0970</td>
<td>.0124</td>
<td>.031</td>
<td>.2241</td>
</tr>
<tr>
<td>Lactations within cows, V(R)</td>
<td>284</td>
<td>.0712</td>
<td>.0206</td>
<td>.0324</td>
<td>.1402</td>
</tr>
<tr>
<td>Tests within lactations, V(W)</td>
<td>2744</td>
<td>.1713</td>
<td>.0310</td>
<td>.0651</td>
<td>.2167</td>
</tr>
</tbody>
</table>

Table 16. Components of covariance from paternal half sibs

<table>
<thead>
<tr>
<th>Source of covariance component between traits x and y</th>
<th>FP</th>
<th>FS</th>
<th>FT</th>
<th>PS</th>
<th>PT</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires within herds, Cov(JxJy)</td>
<td>.0085</td>
<td>.0066</td>
<td>.0323</td>
<td>.0055</td>
<td>.0134</td>
<td>.0165</td>
</tr>
<tr>
<td>Cows within sires, Cov(KxK_y)</td>
<td>.0384</td>
<td>.0420</td>
<td>.1398</td>
<td>.0265</td>
<td>.0662</td>
<td>.0793</td>
</tr>
<tr>
<td>Lactations within cows, Cov(RxR_y)</td>
<td>.0089</td>
<td>.0232</td>
<td>.0942</td>
<td>.0069</td>
<td>.0160</td>
<td>.0508</td>
</tr>
<tr>
<td>Within lactations, Cov(WxW_y)</td>
<td>.0019</td>
<td>.0011</td>
<td>.0179</td>
<td>.0017</td>
<td>.0034</td>
<td>.0061</td>
</tr>
</tbody>
</table>
In Table 15, the variance between sires, $V(J)$, is expected to contain $\frac{1}{2}V(G)$ plus a fraction of any epistatic variance caused by interactions between the additive deviations of non-allelic genes. Letting $m$ designate the number of genes involved in a particular "additive x additive" contribution, $V(A^m)$, the fraction expected in the component between sires is $\sum_{m>1} (\frac{1}{2})^m V(A^m)$. Of course, $V(J)$ also may contain some environmental variance. Even though it is hoped that the sire component has been freed of environmental contributions, one is never absolutely certain that this has been accomplished.

The coefficient of relationship between cows by the same sire was taken as $\frac{1}{2}$ because no important deviations from random mating toward consanguinity were noted in the pedigrees. Except, perhaps, in the case of fat content, assortive mating hardly could have been practiced since the phenotypes of mates were mostly unknown.

The component of variance between cows within sires, $V(K)$, contains the portions of the components; $V(G)$, $V(1)$, $V(D)$ and $V(Z)$ which were not removed in the variance between sires. Definitions for the components between lactations, $V(R)$, and within lactations $V(W)$ are the same as were given previously.

It follows that the variance among cows, each with a single lactation average, is estimated by;
\[ V(\bar{C}) = V(J) + V(K) + V(R) + \frac{V(W)}{9}. \]

The same properties hold for the covariance components in Table 16; hence the covariance between lactation averages of cows for traits x and y;

\[ \text{Cov}(\bar{C}_x \bar{C}_y) = \text{Cov}(J_x J_y) + \text{Cov}(K_x K_y) + \text{Cov}(R_x R_y) + \frac{\text{Cov}(W_x W_y)}{9}. \]

Data from 240 daughter-dam pairs were available for the genetic analysis. These data were adjusted with the constants for herd-month subclasses and with the regressions on ages and on stages of lactation derived in the least-squares analysis.

Some dams had more than one daughter. The pairs were formed from 175 different dams by repeating the dam's record with the record of each of her daughters. To calculate the covariances between daughter and dam, the average test of the lactation with the largest number of tests for each cow was used. The number of tests per cow averaged 4.3. Table 17 shows the pertinent covariances between daughter and dam. These covariances estimate directly what would have been obtained if the lactation averages were, in fact, based on nine tests. This happens because the expectation of the covariance between relatives is independent of the number of observations in each animal's average.

The covariances on the diagonal of Table 17 estimate
\[ \frac{1}{2}V(G) + \sum_{m>1} \left( \frac{1}{2} \right)^m V(A^m) \]

plus the variance from any other effects which make the daughters resemble their own dam more than other dams. For the expectations of the off-diagonal covariances, the appropriate covariance terms would be substituted for the variances in the above formula. If maternal effects are not important, the reciprocal covariances between daughter and dam supposedly estimate the same quantity, for instance, 
\[ \text{Cov}(FP) = \text{Cov}(F'P). \]

Some of the reciprocal covariances in Table 17 deviate considerably from equality, but no explanation for this, other than sampling error, is apparent.

Adjustment of the heritabilities from the daughter-dam pairs to a lactation average basis required computing separately the variances between and within dams. These components of variance are shown in Table 18. The estimates of the phenotypic variances between lactation averages, \( V(\bar{C}) \), were obtained by adding \( 1/9 \) of the variance within dams to the components of variance between dams.

The standard deviations of lactation averages in Table 19 were estimated from the variance components of Table 15 by
\[ \sqrt{V(J) + V(K) + V(R) + \frac{V(W)}{9}}. \]

These estimates agree well with those of .16 to .17 for protein content reported by Politiek (1957), and of .17 to .20 for protein and .26 to .27 for solids-not-fat found by Robertson et al. (1956). As would be expected, the standard deviation of fat content is between
Table 17. Covariances from daughter-dam pairs

<table>
<thead>
<tr>
<th>Dams</th>
<th>Fat,(F)</th>
<th>Protein,(P)</th>
<th>Solids-not-fat,(S)</th>
<th>Total solids,(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat,(F')</td>
<td>.0576</td>
<td>.0232</td>
<td>.0379</td>
<td>.0959</td>
</tr>
<tr>
<td>Protein,(P')</td>
<td>.0082</td>
<td>.0086</td>
<td>.0135</td>
<td>.0217</td>
</tr>
<tr>
<td>Solids-not-fat,(S')</td>
<td>.0098</td>
<td>.0100</td>
<td>.0193</td>
<td>.0283</td>
</tr>
<tr>
<td>Total solids, (T')</td>
<td>.0676</td>
<td>.0333</td>
<td>.0573</td>
<td>.1254</td>
</tr>
</tbody>
</table>

Table 18. Components of variance from dams

<table>
<thead>
<tr>
<th>Source of component</th>
<th>D.f.</th>
<th>Fat</th>
<th>Protein</th>
<th>Solids-not-fat</th>
<th>Total solids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between dams</td>
<td>174</td>
<td>.2019</td>
<td>.0324</td>
<td>.0630</td>
<td>.4172</td>
</tr>
<tr>
<td>Within dams</td>
<td>686</td>
<td>.1553</td>
<td>.0323</td>
<td>.0810</td>
<td>.2400</td>
</tr>
<tr>
<td>Between dams + 1/9(within dams), V(ζ)</td>
<td></td>
<td>.2192</td>
<td>.0360</td>
<td>.0720</td>
<td>.4439</td>
</tr>
</tbody>
</table>
those usually reported for the high and low testing breeds.

The repeatabilities of individual tests in Table 19 are intraclass correlations derived from the components of variance in Table 15, i.e. \( r_1 = \frac{V(J) + V(K) + V(R)}{V(J) + V(K) + V(R) + V(W)} \).

The correlation, \( r_1 \), is the ratio of the component of variance expected between cows, each with a single lactation, to the total variance between individual tests. These repeatabilities can be used to compare the accuracies of different sampling schemes for predicting a cow's true lactation average or breeding value, provided that the schemes are such that the average effects of stage of lactation and month of year are approximately eliminated from the differences between lactations. In other words, these repeatabilities should be valid for comparing, say, semimonthly, monthly and bimonthly testing. Repeatability is higher for total solids than for either of its two constituents because the correlation of fat with solids-not-fat is larger between cows than it is within cows.

The repeatabilities of lactation averages in Table 19 were also derived from the components of variance in Table 15. The appropriate function of the components to estimate the intraclass correlation is, \( r_c = \frac{V(J) + V(K)}{V(J) + V(K) + V(R) + V(W)} \).

The repeatability for protein content is slightly smaller than
Table 19. Intra-herd standard deviations, heritabilities and repeatabilities

<table>
<thead>
<tr>
<th></th>
<th>Fat</th>
<th>Protein</th>
<th>Solids-not-fat</th>
<th>Total solids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviations of lactation averages</td>
<td>.46</td>
<td>.20</td>
<td>.28</td>
<td>.66</td>
</tr>
<tr>
<td>Repeatabilities of individual tests, $r_I$</td>
<td>.53</td>
<td>.54</td>
<td>.53</td>
<td>.66</td>
</tr>
<tr>
<td>Repeatabilities of lactation averages, $r_C$</td>
<td>.58</td>
<td>.41</td>
<td>.52</td>
<td>.62</td>
</tr>
<tr>
<td>Heritabilities from paternal half sibs</td>
<td>.49</td>
<td>.44</td>
<td>.37</td>
<td>.44</td>
</tr>
<tr>
<td>Heritabilities from daughter-dam pairs</td>
<td>.52</td>
<td>.48</td>
<td>.53</td>
<td>.56</td>
</tr>
</tbody>
</table>

The heritability and, since repeatability sets an upper limit on heritability, this is theoretically impossible. However, sampling errors could easily account for the small discrepancy observed here.

The pertinent intraclass correlations were used to compute the heritabilities shown in Table 19 from paternal half sibs. To estimate these heritabilities, the components of variance from Table 15 were substituted in the formula,

$$h^2 = \frac{4V(J)}{V(J) + V(K) + V(R) + V(W)/9}.$$  

Heritabilities from the daughter-dam pairs were estimated by doubling the regressions of daughter on dam derived
from the pertinent covariances in Table 17 and the variances in Table 18; for example, heritability of fat content,

\[ h^2_F = \frac{2 \cdot \text{Cov}(FF')}{V(C_F)} = \frac{2 \cdot 0.0576}{.2192} = .52. \]

Regressions were used, rather than correlations, for computational reasons.

Considering the size of the samples, the heritabilities from paternal half sibs and daughter-dam pairs agree remarkably well. Because about half of the cows have records in both sets of data some automatic resemblance exists, but the estimates from the two sources should be nearly independent. This is because one heritability is based on the resemblance between a cow and her paternal sister while the other is based on her resemblance to her daughter or to her dam.

It was impossible to calculate exact standard errors for the statistics. However, by performing some rough approximations it appeared that the heritabilities and genetic correlations from the two sources should be about equally reliable in these data. Like the heritabilities, the genetic correlations are also nearly equal for the daughter-dam pairs and paternal half sibs. Consequently, it appeared reasonable to take a simple average of the estimates from the two sources as the combined estimate of each parameter.

If the data had been, in fact, orthogonal, the standard error for the heritabilities would have been of the order of
.14 for both the daughter-dam pairs and the paternal half sibs. This would establish a lower limit of about .10 on the standard errors of the pooled heritabilities, but the amount which the actual standard errors are larger than .10 is unknown.

The heritabilities of fat content in Table 19 are very close to the values of .5 to .6 commonly reported for this trait. Averaging the heritabilities for solids-not-fat content from the two sources gives an estimate only slightly larger than that reported by Johnson (1957). The average heritabilities for both protein and solids-not-fat are approximately equal to those found by Robertson et al. (1956) and Politiek (1956). The estimates in Table 19 certainly do not contradict the hypothesis that the true heritabilities of these milk constituents are of the order of .5.

The genetic correlations are presented in Table 20. The genetic correlations for traits x and y from the paternal half sibs were computed from the appropriate sire components in Tables 15 and 16 by the formula, 
\[ r_{GxGy} = \frac{Cov(JxJy)}{\sqrt{V(Jx)V(Jy)}}. \]
For example, the genetic correlation between fat and protein from the paternal half sibs was computed by 
\[ \frac{.0085}{\sqrt{(0.0260)(0.0045)}} = .78. \]
The corresponding estimate from the daughter-dam pairs was derived from Table 17 by taking, 
\[ \sqrt{\frac{Cov(FP')} {Cov(FF')} \frac{Cov(F'P)} {Cov(PP')}} = .62. \]
### Table 20. Intra-herd genetic and phenotypic correlations

<table>
<thead>
<tr>
<th></th>
<th>FP</th>
<th>FS</th>
<th>FT</th>
<th>PS</th>
<th>PT</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic correlations from</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paternal half sibs</td>
<td>.78</td>
<td>.53</td>
<td>.91</td>
<td>.94</td>
<td>.91</td>
<td>.87</td>
</tr>
<tr>
<td>Genetic correlations from</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>daughter-dam pairs</td>
<td>.62</td>
<td>.58</td>
<td>.95</td>
<td>.90</td>
<td>.82</td>
<td>.82</td>
</tr>
<tr>
<td>Phenotypic correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>between lactation averages</td>
<td>.62</td>
<td>.54</td>
<td>.81</td>
<td>.70</td>
<td>.74</td>
<td>.81</td>
</tr>
</tbody>
</table>

In those cases where the correlation is between a constituent and one of its component parts, the correlations are large partly because of the part-whole relationship. For example, the genetic correlation between fat and total solids is expected to be, \( r_{GF \cdot GT} = \sqrt{\frac{V(G_F)}{V(G_T)}} + r_{GF GS} \sqrt{\frac{V(G_S)}{V(G_T)}} \). Thus, even if the genetic correlation between fat and solids-not-fat were zero, \( r_{GF GT} = \sqrt{\frac{V(G_F)}{V(G_F) + V(G_S)}} \).

Except for the average genetic correlation between fat and protein being about .2 higher in these data than in those of Robertson et al. (1956), the genetic correlations are substantially the same in the two studies. The large genetic correlations of fat with both protein and solids-not-fat certainly contradict the statement often seen in the literature to the effect that protein and/or solids-not-fat are inherited to a "large degree" independently of fat.

Any genetic correlation between two traits could be
caused by the primary effects of the same genes acting on the two traits or by some genetically controlled common cause. For example, the genetic mechanism causing the large positive correlation between fat and solids-not-fat could be that all cows secrete about equal quantities of these constituents, but some cows possess genes which cause them to put more water in their milk than other cows. Unfortunately when one can only observe the phenotypes of the traits, as with quantitative characters, no way is apparent for separating the possible causes of genetic correlations.

For estimating the phenotypic correlations between lactation averages in Table 20, the components of variance and covariance in Tables 15 and 16 were combined into the previously defined estimates of $\text{Cov}(\overline{C}_x, \overline{C}_y)$, $V(\overline{C}_x)$ and $V(\overline{C}_y)$. These phenotypic correlations fall well within the range reported by other workers.

Some of the regression coefficients between lactation averages are also of interest. Estimating the regressions from the components in Tables 15 and 16, the regression of solids-not-fat on fat is .33 and of protein on fat is .27. The regression of solids-not-fat on protein is .99.

In order to inspect some of the more important relations between the lactose-ash portion and the other constituents, statistics were derived from the components of variance and
covariance between the paternal half sibs. The derivations were accomplished by estimating the pertinent components from Tables 15 and 16 as described for the analysis of individual samples, i.e. $V(Q) = V(S-P)$ etc. Then these components were combined as described for the heritabilities, and genetic and phenotypic correlations of the other constituents.

The heritability of lactose-ash is .07. The genetic correlations involving lactose-ash are -.37 with fat, .48 with protein and .74 with solids-not-fat. The corresponding phenotypic correlations are .12, .00, and .62 respectively. Except for the lower heritability and the negative genetic correlation between fat and lactose-ash, these estimates agree as well as could be expected with those given for lactose by Robertson et al. (1956).
DISCUSSION

Although some of the statistical methods used in this study were only approximate, no serious consequences of this were indicated in those cases where the present results are directly comparable to the results of other investigations. Of course, the application of statistics to actual data nearly always involves approximate methods to some extent.

The analysis of factors affecting the composition of individual samples of cow's milk disclosed some interesting relations. Among these is the considerable effect of age on the solids-not-fat. This remains even after accounting for the effects of the higher frequencies of damaged udder tissue and mastitis in older cows. Perhaps this results from the udder tissue becoming generally more permeable to the blood constituents as the cows age and, thus, requiring less lactose content to maintain the osmotic equilibrium between blood and milk.

If the age component for solids-not-fat in Table 5 is added to the expected variance between lactation averages derived from Table 15 and then expressed as a percentage of the sum of the two, 13 per cent is the portion of the total variance between lactation averages of cows not the same age which is due to ages. Similar manipulations for the protein and fat contents yield estimates of three and four per cent
respectively for the variance due to ages.

The classification according to abnormality was dichotomous and, if the criteria used for the separation had been more stringent, larger differences between the normals and abnormals may have occurred. However, the frequency of abnormals in these data (28 per cent) does not seem unreasonably high compared to that ordinarily found in surveys performed especially to estimate the frequency of mastitis.

The only interaction, other than those between herds and months, which seemed to approach practical significance was between herds and stages of lactation in the case of fat content. Other workers have found differences in the lactation curves for fat content from breed to breed. Breed differences may have caused the interaction in these data.

The variances for the interactions between herds and months could have been divided into components for areas by months and for herds within areas by months, similar to the division for breeds shown in Table 7. As indicated by the larger differences between the herd-month constants for the winter and summer months in area III than in area I for both protein and solids-not-fat, part of these herd-month interactions may have been caused by the herds being in different areas. However, even if this were true, one could not separate the climatic and managemental effects in these data.
Further experimentation is needed if the possible causes of the interactions between months and herds are to be clarified.

That the increase in protein content soon after the cows are first put to pasture in the spring noted by European workers did not occur in the present data, perhaps resulted from a higher nutritional level being maintained in these herds throughout the winter. In general, the herds in this study were fed according to or in excess of the recommended standards.

The relative magnitudes of the components of variance for protein and lactose-ash, and the correlations between these constituents and solids-not-fat in Tables 8 and 9, indicate that the temporary environmental variance of solids-not-fat is caused largely by variation in the lactose, but the genetically caused variation in solids-not-fat is controlled mostly by the protein. The variation which physiological factors cause in solids-not-fat apparently comes largely from differences in protein for stage of lactation and from differences in lactose for ages. Except for the components for age and for stage of lactation, the variance in total solids seems to be caused somewhat more by differences in the fat content than by differences in solids-not-fat.

The large correlations among the constituents for breeds indicate positive genetic correlations between the constitu-
ents since it would seem to be a coincidence if the higher contents of solids-not-fat and protein occurred with higher fat content merely because of "genetic drift" (chance) rather than through correlated responses to selection.

The regressions of solids-not-fat on fat for each source given in Table 10 show that there is no hope of finding a simple prediction equation which will be reasonably accurate under all conditions. However, if one is forced to predict the content of the non-fat solids from the fat, and little or nothing is known about the origin of the milk, the accepted equations with the regression coefficient in the vicinity of .4 can be used. Of course, no account has been taken here of the possibility of curvilinearity in the relation between these two constituents.

Ordinarily, the lactation average for fat content would be formed by weighting each test by the milk production during the corresponding month. In this study the milk yields were not recorded for some of the composites; hence the lactation average was defined as an average of equally weighted tests. The means would be slightly lower for the weighted tests, but the only concern here is the possibility of differences in variances. The tests during the last two months of lactation are known to be more variable than those in the middle months. Because cows produce less milk during the last
two months of lactation, these tests would contribute less to the weighted average than they would to the simple average. Though it is not known exactly, it is thought that any differences in the variance between lactation averages from weighted and from unweighted tests probably would be of little practical importance.

If one is satisfied with the accuracy obtained by monthly testing for fat content, the near equality of the repeatabilities of individual tests in Table 19 indicate that monthly testing is also appropriate for protein or solids-not-fat. This conclusion disagrees with that of some workers who have reasoned (fallaciously), that because the environmental variance is smaller for protein than for fat content, less frequent testing is required for protein to obtain equal accuracy.

Assuming that a testing scheme is carried out so as to eliminate approximately the "fixed" effects of stage of lactation and month of year in the differences between cows with single lactation averages, the formula given by Lush (1948) can be used to estimate the relative gain or loss in progress per generation by selecting on single lactation averages which are based on different numbers of tests. Where m number of tests is compared to n tests this formula is \( \sqrt{\frac{h^2_n}{h^2_m}} \) or
\[
\sqrt{\frac{m}{n} \frac{1 + (n-1)r^I}{1 + (m-1)r^I}}
\]
in terms of repeatability of individual tests. If bimonthly testing is compared to monthly testing; say, \( m = 5 \), \( n = 9 \) and \( r^I = .5 \), progress per generation for fat, protein or solids-not-fat contents would be approximately \( \sqrt{\frac{25}{27}} = .96 \) times as much by changing to bimonthly testing. By going to a semimonthly testing scheme \( (m = 18) \), progress is expected to be approximately 1.026 times faster than would be obtained with monthly testing. Even though these estimates are only approximate, they indicate that bimonthly testing would be satisfactory, especially when the cost of making each test is high.

The repeatabilities of the lactation averages themselves, from lactation to lactation, are perhaps a bit lower than expected. If the true values of repeatability and heritability actually are as nearly alike for each constituent as the estimates were in these data, this would leave little room for dominance and epistatic deviations or for permanent environmental effects to cause differences between cows. That fat content is not influenced much by dominance deviations is already indicated by the average fat per cent remaining relatively constant under inbreeding, and not showing much if any heterosis in first crosses in crossbreeding experiments.

It seems well established that the heritabilities of
protein and solids-not-fat contents are high enough for mass selection to be effective in changing the population means for these traits. The genetic progress per generation, $\Delta G$, is expected to be the product of heritability and the selection differential, i.e. $\Delta G = h^2(d - \bar{d})$. The selection differential, $d - \bar{d}$, is the difference between the mean of those animals selected to be parents of the next generation and the mean of the population from which they were selected.

For protein and non-fat solids, heritability is high and one would obtain in the progeny much of the difference which was between $d$ and $\bar{d}$ in the parental population, but progress would be relatively slow because the small standard deviations of these traits signify that $d - \bar{d}$ can not be large. Although information on progeny or collateral relatives would be useful if available at the time selections were made, waiting for this information after an estimate of a cow's own phenotype is available would decrease progress per year.

When heritability is high, a progeny test on a sire can be made relatively accurate with only a few daughters. Provided that the environmental correlation between daughters of the same sire is zero, the correlation between a sire's breeding value, $G$, and the average phenotypes of $n$ of his offspring, $\bar{P}$, in a random mating population is, $r_{GP} = \sqrt{\frac{nh^2}{4(n-1)h^2}}$. 
Lush (1931). If heritability is .5, \( r_{GP} \) is already .65 when \( n = 5 \), rises to .77 for \( n = 10 \) and to .86 for \( n = 20 \). Using statistics of approximately the same magnitudes as found in this study, Robertson et al. (1956) estimated that culling the lowest third of the first lactation cows on solids-not-fat content and selecting sires from the highest 5 per cent of the cows would lead to an average yearly increase of .02 per cent in non-fat solids. Similar arguments for selecting wholly on protein content yield an expected yearly increase of about .012 per cent. These are estimates of the maximum average changes because culling a third of the cows expends approximately all of the freedom to cull.

For predicting the correlated response in fat content when the selections are based wholly on the solids-not-fat, one needs the genetic regression of fat on solids-not-fat, i.e.

\[
b_{GF}G_S = \frac{\text{Cov}(GF GS)}{\text{Var}(GS)} = r_{GF}G_S \sqrt{\frac{\text{Var}(GF)}{\text{Var}(GS)}}
\]

Then \( \Delta G_F = \Delta G_S (b_{GF}G_S) \), and direct selection for solids-not-fat content intense enough to increase its mean 1.0 is expected to cause an increase of \( (b_{GF}G_S) \) in fat content. Using the averages of the genetic correlations and heritabilities from the paternal half sibs and daughter-dam pairs, \( b_{GF}G_S = .56 \sqrt{\frac{.105}{.035}} = .97 \). Table 21 shows these genetic regressions and the expected minimum time required to change the constituent on which selection is based. In 19 years of selecting only for total solids content, its mean is expected to change
1.0 and protein, solids-not-fat and fat are expected to change .25, .34 and .64 respectively. Table 22 shows these estimates expressed on the basis of equal total selection time.

A selection index such as recently proposed by Kempthorne and Nordskog (1959) could be used to select for protein and/or solids-not-fat with the restriction that the fat content remain constant. However, this would reduce further the small changes expected from selecting solely on protein or solids-not-fat.

The average genetic correlation of .92 between protein, as measured by formol titration, and solids-not-fat indicates that formol titration would be a satisfactory field test for measuring the non-fat solids. Comparing the results for protein in this study with those of other workers, who used the Kjeldahl method indicates that formol titration is a reliable test for total protein.

If the solids-not-fat had been estimated with a prediction equation using fat and lactometer readings as the "independent" variables, the genetic correlation between fat and solids-not-fat would be expected to be larger than found in this study. This is because fat and specific gravity are given approximately equal weight in the prediction equation and the genetic correlation is that between fat and fat plus specific gravity.
Table 21. Expected average change in constituent Y, when selection is based wholly on constituent X and changes X by 1.0%.

<table>
<thead>
<tr>
<th>X</th>
<th>Protein</th>
<th>Solids-not-fat</th>
<th>Fat</th>
<th>Total solids</th>
<th>Time required (Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>1.00</td>
<td>1.29</td>
<td>1.69</td>
<td>3.01</td>
<td>67</td>
</tr>
<tr>
<td>Solids-not-fat</td>
<td>.66</td>
<td>1.00</td>
<td>.97</td>
<td>2.01</td>
<td>50</td>
</tr>
<tr>
<td>Fat</td>
<td>.29</td>
<td>.32</td>
<td>1.00</td>
<td>1.35</td>
<td>30</td>
</tr>
<tr>
<td>Total solids</td>
<td>.25</td>
<td>.34</td>
<td>.64</td>
<td>1.00</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 22. Expected average change in constituent Y when selection is based wholly on constituent X.

<table>
<thead>
<tr>
<th>X</th>
<th>Protein</th>
<th>Solids-not-fat</th>
<th>Fat</th>
<th>Total solids</th>
<th>Time required (Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>.28</td>
<td>.37</td>
<td>.48</td>
<td>.85</td>
<td>19</td>
</tr>
<tr>
<td>Solids-not-fat</td>
<td>.25</td>
<td>.38</td>
<td>.37</td>
<td>.80</td>
<td>19</td>
</tr>
<tr>
<td>Fat</td>
<td>.18</td>
<td>.20</td>
<td>.64</td>
<td>.85</td>
<td>19</td>
</tr>
<tr>
<td>Total solids</td>
<td>.25</td>
<td>.34</td>
<td>.64</td>
<td>1.00</td>
<td>19</td>
</tr>
</tbody>
</table>
SUMMARY

Data collected from 23 well managed herds of the five major dairy breeds, located in the principal dairying areas of Oregon, were used to estimate the importance of factors affecting the composition of individual samples of cows' milk and to derive genetic and phenotypic statistics concerning the lactation averages. Each farm was visited about nine times, as an average. The average interval between visits was approximately six weeks.

Each individual milk sample was classified as normal or abnormal for mastitis. The daily composites were analyzed for total solids by the Mojonnier procedure, for fat by the Babcock test and for protein by formol titration. Solids-not-fat were determined by difference. Occasionally, the estimate by difference of lactose-ash was used.

To circumvent the problem of heterogeneous variances in pooling the data from the five breeds, square root transformations, as indicated by the data themselves, were made. However, the estimates from the transformed and from the original data didn't differ appreciably.

From 4,462 observations on each constituent, components of variance and covariance were estimated for the factors affecting the composition of the individual samples. The com-
ponent for normal versus abnormal milk was nearly zero for each constituent. All of the two-factor interaction components among herds, months, stages of lactation and ages were estimated and, except for the herd by month components for protein and solids-not-fat, none contributed more than 2.7 per cent to the total variation of a constituent.

Differences between breeds contributed from 50 to 65 per cent to the total variation of each constituent, except for lactose-ash where it was only four per cent. The components for herds within breeds ranged from one per cent of the total variation for protein to five per cent for lactose-ash. Differences between months accounted for four per cent of the total variation in fat and total solids, two per cent for solids-not-fat and lactose-ash and was nearly zero for protein. The components for lactation stages were moderately large for all the constituents except lactose-ash. Differences between ages caused three per cent of the total variation of individual samples in solids-not-fat and lactose-ash, and one per cent in the other constituents. The components for interaction between months and herds within breeds contributed 25, 8 and 5 per cent to the total variance of lactose-ash, protein and solids-not-fat respectively. The components for cows within herds ranged from 11 to 15 per cent of the total variance of each constituent.
The components of variance and covariance were used to calculate the correlations among the constituents for each source of variation. The correlations among fat, protein, solids-not-fat and total solids were all positive and large for breeds, herds within breeds, months, stages of lactation, ages and cows within herds. The correlations among fat, protein and solids-not-fat were all small for the herd by month interaction and error deviations, and protein was negatively correlated with lactose-ash for these two sources.

Path coefficients were used to combine the correlations between fat and solids-not-fat for the different sources into the total, intra-breed and intra-herd correlations between these two constituents.

The method of "fitting constants" was used to estimate the effects of ages, stages of lactation and the herd-month sub-classes. Fat, protein, solids-not-fat and total solids were all lowest during the second month of lactation and increased steadily until the lactation terminated. The effects of age, although small, seemed to be curvilinear for fat, protein, and total solids. Solids-not-fat and lactose-ash decreased in an essentially linear manner as age increased. The constituents were generally lower during the summer than during the winter months.

The effects of herds, months, herd by month interactions,
ages and stages of lactation were removed for the genetic analysis. Lactation averages were calculated as those expected from nine equally weighted tests. The standard deviations of lactation averages were .46, .20, .28 and .66 for fat, protein, solids-not-fat and total solids, respectively.

The repeatabilities of individual tests were approximately .5 for fat, protein and solids-not-fat, indicating that testing should be equally frequent for these constituents to obtain the same accuracy in predicting a cow's true lactation average. The repeatabilities of lactation averages were approximately .4 for protein, .5 for solids-not-fat and .6 for fat and total solids.

Heritabilities and genetic correlations were calculated from 240 daughter-dam pairs and from 232 paternal half sister families totaling 793 cows. The average heritabilities from the two kinds of relatives were of the order of .50 for fat and total solids and .45 for protein and solids-not-fat. Genetically, fat was correlated with protein .70, with solids-not-fat .55 and with total solids .93. The average genetic correlation between protein and solids-not-fat was .92, and between solids-not-fat and total solids it was .85. The last three are partly automatic. The phenotypic correlations using the lactation average as the basic unit, were all slightly smaller than the genetic correlations.
Heritability of lactose-ash, from the paternal half sibs only, was .07. The genetic correlations between lactose-ash and the other constituents were -.37 with fat, .48 with protein and .74 with solids-not-fat. The last would contain some automaticity, but not much because the additive genetic variance of lactose-ash is not large.


—. 1948. The genetics of populations. Ames, Iowa, (Mimeographed.)


ACKNOWLEDGMENTS

It is a pleasure to acknowledge the friendly counsel and encouragement of Dr. Jay L. Lush and Professor O. Kempthorne during the preparation of this thesis and throughout the author's graduate study.

Appreciation also is expressed to Dr. G. A. Richardson, Professor J. O. Young and Professor P. M. Brandt, of the Department of Dairy Husbandry at Oregon State College, for making available the data used in this study.