17244

A STUDY OF WITHIN HERD VARIABILITY IN MILK FAT, PROTEIN AND LACTOSE CONTENT OF BULK MILKS IN BRITISH COLUMBIA AND FACTORS AFFECTING THE DESIGN OF HERD MILK SAMPLING PROGRAMS

by

CHRISTOPHER JOHN WILLIAMS

B.S.A., University of British Columbia, 1967
M.Sc., University of British Columbia, 1971

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in the Department

of

Animal Science

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May, 1973

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Animal Science

The University of British Columbia
Vancouver 8, Canada

Date May 10 1973

ABSTRACT

Three sets of data were used to estimate variation, from all sources, associated with bulk milk sampling and testing programs. Three milk samples were taken from each shipment of 26 herds from March 14, 1970 to April 24, 1971 (Experiment I). The set of three samples was handled as follows: (1) one sample was used in the formation of a two-week composite; (2) one sample was used in the formation of a one-week composite; and (3) one sample was analysed fresh. Four milk samples were taken from each shipment of 22 different herds from November 17 to December 16, 1971. Three of the four samples were analysed fresh in duplicate (Experiment II). The fourth sample was divided into three parts and each part was used in the formation of a composite. Each composite was analysed in duplicate after a two-week collection period (Experiment III). Herd milk was shipped on alternate days. All milk samples (8,894) were analysed for milk fat, protein and lactose using Infrared Milk Analysers.

Estimates, obtained from Experiment I by the analyses of variance of a hierarchal model (herds, periods within herds and shipments within herds and periods), of within herd-period (15 shipments per period) variances of percent

milk fat, protein and lactose were; 0.01371 ± .00030, 0.00787 ± .00017 and 0.00548 ± .00012 respectively. Estimates were obtained from Experiment II of within herd-period variance and its components by the analyses of variance of a hierarchial model. The estimates of these variances for percent milk fat, protein and lactose respectively were: (1) within herd-period variance -- 0.01329 ± .00064, 0.00507 ± .00031 and 0.00483 ± .00017; (2) biological (shipment to shipment) variance -- 0.00607 ± .00061, 0.00340 ± .00029 and 0.00110 ± .00014; (3) sampling (within shipment) variance 0.00094 ± .00027, -.00021 ± .00006 and -.00033 ± .00013; and (4) testing (within sample) variance -- 0.00628 ± .00029, 0.00167 ± 0.00006 and 0.00373 ± .00013. Estimates of within herd-period variance of percent protein from Experiment I were significantly different from estimates from Experiment II.

Orthogonal polynomials were used to estimate the relationship between the serial correlations (calculated from Experiment I) of milk constituent percentage and the number of shipments separating two shipments for which the correlations were calculated. Only the linear term was significant for percent protein and lactose and accounted for 99.7 and 98.4 percent of the total sums of squares for these two milk constituents respectively. Linear and quadratic after linear were significant for percent milk fat serial correlations and accounted for 98.4 and 1.3

percent of the total sums of squares respectively.

Strata within periods was fitted as an effect (Experiments I and II) in a hierarchal model and was a significant source of variation. The variances of estimates of herd-period mean milk constituent percentages obtained from various simple and stratified random sampling schemes were calculated. Stratification resulted in a relatively small reduction in the variances of these estimates.

Estimates of the variances associated with the formation of a composite sample obtained from Experiment III by the analysis of variance and from Experiments I and II were near zero. The variance of estimates of herd-period mean milk constituent percentages obtained from two two-week composites were 0.00368, 0.00110 and 0.00205 for percent milk fat, protein and lactose respectively. It was calculated that four random samples would estimate herd-period mean milk constituent percentages at least as precisely as two two-week composite samples.

Two-week composite samples underestimated percent milk fat by 0.045 percent milk fat and overestimated percent protein and lactose by 0.023 and 0.010 percent respectively compared to corresponding estimates based on the fresh analyses of samples drawn from each shipment.

Simple and multiple regression techniques were used in an attempt to predict herd differences in within herd-period variance from the average amount of milk shipped and percent milk fat, protein and lactose. In general, large within herd-period variances of milk constituent percentages were significantly associated with small herd milk shipments and high levels of milk fat and protein. However, the proportion of the total sums of squares accounted for by the various regression equations was relatively low; therefore the equations were not useful for predicting herd-period variances.

Within herd-period variance of percent milk fat was highest in the spring and autumn; therefore sampling frequency may need to be greater at some seasons than at others. Differences among herds in within herd-period variance of milk constituent percentages were significant; therefore random sampling schemes may have to be modified to suit individual herds.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

ACKNOWLEDGEMENTS

The author wishes to thank Dr. R.G. Peterson, under whose supervision this study was conducted, for his assistance in planning the project and in analysing the results. The author also thanks Dr. C.W. Roberts and Dr. J. Hodges for their suggestions and criticisms. Thanks are extended to the personnel of the British Columbia Department of Agriculture, Dairy Branch: Mr. T.C.T. Chao, Technical Director; Mr. G.D. Johnson, Officer-in-Charge, Dairy Laboratory; and Mr. E.N. Jenstad, Dairy Specialist for supervision of data collection and for the analyses of milk samples. The author also expresses his sincere thanks to his wife for her support and encouragement.

INTRODUCTION

The producer price for whole milk is usually established per hundred pounds of milk of a given milk fat (and/or other milk constituent) percentage. This basic price is adjusted for deviations of the milk shipped by individual dairy farmers from the given percentage. Therefore, determining the percent composition of herd milk is important in paying producers accurately.

At the present time, British Columbia producer milk prices are established each month by a pricing formula which includes a differential for deviations from the given percentage for milk fat only. The accounting period, in British Columbia, is a calendar month and it is necessary to sample herd milk in order to determine the monthly average percent milk fat.

In general three sampling schemes can be proposed: (1) drawing a sample from each shipment and forming a composite which is tested after a collection period of several days (composite sampling); (2) drawing a milk sample from each shipment and testing the sample fresh (fresh sampling) and (3) drawing a milk sample from randomly selected shipments and testing fresh (random sampling). Other methods are also possible but have serious drawbacks, for example:

1. systematic selection of shipments - such as sampling every fourth or fifth shipment--can lead to biased estimates;
2. formation of a composite of samples from randomly selected shipments includes the disadvantages inherent in both composite and random sampling.

The first scheme is currently used in British Columbia and the usual compositing period is two weeks. The main disadvantages of composite sampling lie in the labor required to sample each shipment and to transfer the sample to a composite bottle. In addition the compositing procedures and storage of the composites could introduce bias and/or variation in the test results.

The second scheme removes the need for forming and maintaining a composite bottle for each herd but it requires the same number of samples as the compositing method and more laboratory analyses. However, it is the most precise of the three schemes. The third scheme also removes the need for compositing, it requires fewer samples than either of the first two methods and fewer laboratory analyses than the second scheme but will probably require more laboratory analyses than the first method if it is to be as precise. However, the advent of automatic milk analysers has reduced the time and costs of milk analyses. This equipment can output the test results on punched tape and thus facilitate computer handling of test information. The main costs of bulk milk sampling are due to the collection and handling of milk samples. Estimates of herd-period means from random samples contain variation due to true differences between shipments; this is not a source of variation in estimates obtained from either of the first two schemes. Therefore random sampling can not be as precise as the second method but should yield unbiased estimates of the true herd-period percent milk composition.

The purpose of this study was to estimate the variability, from all sources, of estimates of percentages of milk fat, protein and lactose in bulk tank milk shipments and to consider ways of assessing the percent milk fat, protein and lactose in herd milk without composite samples. Only a small amount of research has been done on sampling and testing bulk tank milk and there is a need for a thorough analysis of all sources of variation based on more comprehensive data and longer time periods than has been done in most reported studies. Estimates of the variances associated with bulk tank milk sampling and testing are needed if the precision of various sampling schemes is to be compared.

LITERATURE REVIEW

In studies of the variability of milk fat and total solids content of bulk herd milk in Scotland, O'Keeffe [15,16] sampled each daily bulk tank shipment of ten herds for twelve months. The milk samples were analysed for percent milk fat by the Gerber method and for percent total solids by the Claesson milk testing machine. From these data he estimated the between daily shipment within herd-month variances as 0.0246 and 0.039 for percent milk fat and percent total solids respectively; estimates of the within herd-year variances were 0.043 and 0.085 for the same two milk components respectively. Morris et al. [14] took biweekly milk samples from bulk milk shipments of 88 Minnesota herds for one year. Their estimates of the within herd-year standard deviations were; 0.227 for percent milk fat, 0.181 for percent protein and 0.147 for percent solids-not-fat.

Edwards and Donaldson [7] sampled daily bulk tank milk shipments of thirty-two British herds for thirteen days. The milk samples were analysed for percent milk fat by the Gerber method and for total solids by a gravimeter method. The solids-not-fat percentage was calculated by difference. Their estimates of the between shipment within herd variances for the thirteen day period were 0.0227, 0.0114, and 0.0235 for percent milk fat, solids-not-fat and total solids respectively. These workers computed the difference between consecutive shipment tests and found that the majority of day-to-day differences were small, 83 percent of the differences were less than 0.19 percent milk fat, but the largest difference was 0.63 percent milk fat. In O'Keefee's [16] study the largest day-to-day difference was 1.0 percent milk fat. Edwards and Donaldson [7] reported that the 95 percent confidence interval of the difference between two randomly selected single milk samples was ±.39 percent milk fat. These workers found that while there was a tendency for small herds to have greater between shipment variation than large herds the differences among herds were not significant ($p \leq .05$) by the analysis of variance when the herds were placed into three groups on the basis of the amount of milk shipped.

Herrmann and Anderson [11] in a comprehensive study of milk fat testing in the U.S.A., sampled 49,117 milk shipments (two days production in each shipment) from herds shipping to eleven different milk plants over a period of four months for most herds and over a one year period for the remaining herds. The milk samples were tested for percent fat by the Babcock method. These workers estimated that the within herd-month standard deviation of percent milk fat was 0.146. Boswell et al. [3] sampled daily milk shipments from 86 herds throughout England and Wales for a period of one year. The milk samples were analysed for percent milk fat, solids-not-fat and total solids; the average within herd-month standard deviations of these milk constituents were 0.16%, 0.081% and 0.20% respectively.

Herrmann and Anderson [11] and Boswell et al. [3] found that the within herd-month standard deviation of percent milk fat was highest in November; these two studies reported values of 0.177 percent and 0.19 percent respectively for this month. The lowest standard deviations occurred in the late winter and early spring in both studies; Herrmann and Anderson [11] reported February to be the lowest month (0.137%), while Boswell et al. [3] found March to be the lowest month (0.14%). The study of Boswell et al. [3] showed a secondary peak in May (0.17%). O'Keeffe's [16] study showed the highest within herd-month variance of milk fat percent in May (0.0458) with a second peak in October (0.0381); the lowest values occurred in the winter months with January (0.0085) being the lowest.

Boswell et al. [3] reported that high within herd-month variance of percent milk fat was associated with small herds. Herrmann and Anderson [11] used multiple regression techniques to estimate the effects of: (1) level of milk fat, (2) amount of milk shipped, (3) the coefficient of variation of the amount of milk shipped and (4) the variance of environmental temperature on the within herd-month variance of percent milk fat. The regression model accounted for a significant ($p \leq .05$) reduction in the total sums of squares. Of the four independent variables used only the variance of environmental temperature was not a significant ($p \leq .05$) source of variation. The remaining independent variables were negatively associated with the herd-month variance of percent milk fat.

Herrmann and Anderson [11] found that composite milk samples underestimated percent milk fat as compared to the percentage calculated from fresh milk samples. The average amount of bias was -.011 percent milk fat but varied from -.095 percent to 0.031 percent by milk plant; thus indicating that the amount of bias in composite samples depended on the handling of the samples. Preston [17] also reported that the percent milk fat estimated from composite samples was lower that the corresponding percentage calculated from fresh samples.

To estimate the components of within herd-period variance, O'Keeffe [16] drew triplicate milk samples from the daily bulk shipments of eight herds for eight days. The milk samples were analysed in duplicate for percent milk fat by the Milko-Tester Mark II and for percent total solids by the Claesson milk testing machine. He estimated the variances associated with sampling the bulk tank were 0.0017 and 0.0032 for percent milk fat and total solids respectively. The variances associated with testing were 0.0010 and 0.0024 for percent of the same two milk components respectively; however, O'Keeffe [16] suggested that the testing variance of percent milk fat estimated in this study was much lower than is usually encountered under practical conditions. In a study of bulk tank sampling methods, Dimick and Atherton [5] reported that bulk cooled milk was thoroughly mixed after three minutes of agitation and therefore that the variance associated with sampling bulk tanks is generally low if sampling procedures are carefully followed; these results are supported by Liska et al. [13]. In a review of automatic testing of milk for fat and protein Green [10] reported estimates of the standard deviations associated with testing milk samples on Infrared milk analysers (IRMA) under practical laboratory conditions ranging from 0.06 to 0.09 for both percent milk fat and percent protein. Biggs [2] reported that the standard deviation between duplicates on IRMA equipment was 0.03 or less for all three milk components.

PART 1

ESTIMATION OF POPULATION PARAMETERS

1. INTRODUCTION

The design of a sampling scheme to meet a specified precision requires knowledge of the appropriate population variances. The purpose of Part 1 of this thesis was to estimate the variances associated with sampling bulk tank milk shipments under various sampling schemes. Two main sources of variation are assumed; (1) variation between the true percent composition of shipments (i.e. sampling variance in the statistical sense) and (2) variation associated with the various procedures of estimation. These estimates are used to predict standard errors of herd-period mean milk constituent percentages estimated under different sampling schemes and to determine the number of samples needed if estimates obtained by random sampling are to equal the precision of estimates obtained by composite sampling.

1. MATERIALS AND METHODS

Collection and Analyses of Milk Samples

Three sets of data were collected for this study by drawing samples from bulk milk shipments of Fraser Valley dairy herds. As herd milk was shipped on alternate days, each sample represented two days herd milk production (four milkings). All milk samples were analysed by the British Columbia Department of Agriculture (BCDA), Dairy Branch Laboratory for percent milk fat, protein and lactose using an Infrared Milk Analyser (IRMA). Milk samples were taken by regular Tank Milk Receivers who used the following procedure; bulk milk was agitated for five minutes and then a 100 ml. sample was drawn by taking 20 ml. of milk from each corner and from the middle of each tank. This procedure conforms to the regulations governing sampling of bulk milk and is supposed to be followed by all Tank Milk Receivers when drawing a milk sample. The weight of milk in the shipment was recorded at the time of sampling. The samples were maintained on ice until received at the milk plant.

Experiment I. Three milk samples were taken from each shipment of twenty-six herds, all shipping to the same milk plant, for a period of approximately thirteen months (March 14, 1970 to April 24, 1971). Three herds stopped shipping during the experimental period. The set of three milk samples was handled as follows:

1. one sample was used in the formation of a two-week composite of seven fresh samples;
2. the second sample was used in the formation of a one-week composite of either three or four fresh samples;
3. the third sample was analysed fresh.

Mercuric chloride and potassium dichromate were used as perservatives for the composite samples. The composites were formed in the plant receiving the milk. The total numbers of samples analysed were; 4,701 fresh, 697 two-week and 1,334 one-week.

Experiment II. Four milk samples were taken from each bulk shipment of twenty-two different herds, all shipping to the same milk plant (a different plant than the herds in Experiment I), for a period of one month, from November 17 to December 16, 1971. Fifteen shipments were sampled per herd. Three of the four samples were analysed fresh in duplicate, with duplicates randomly assigned to analysers (1,910 analyses). The fourth sample was used in Experiment III.

Experiment III. The fourth sample collected in Experiment II was divided into three parts and each part was used in the formation of a composite. Each composite was analysed in duplicate after a two-week collection period (252 analyses). Potassium dichromate was used as a preservative for these composites. This set of composites was accumulated in the BCDA, Dairy Branch Laboratory.

The total number of observations for all three experiments was 8,894.

The Problem and Definition of Terms Used

The purpose of this study was to design a random sampling scheme to estimate, with a level of precision acceptable to the dairy industry, the percent milk fat, protein and lactose in milk shipped by farmers during an accounting period. Accounting periods in British Columbia are currently one month long and milk is usually shipped on alternate days therefore a period in this study (unless otherwise specified) was defined as fifteen consecutive shipments. Each herd-period of fifteen shipments was considered to be a finite population of shipments drawn from an infinite population of such herd-period populations.

The word "sample" was used both in its statistical sense and also to refer to a small quantity of milk removed from a shipment of milk for analysis. The meaning intended should be clear from the context in which the word was used. The term "milk constituent" was used to refer to the three main milk constituents (milk fat, protein and lactose) only.

The precision of an estimate may be considered acceptable if differences between estimates of herd-period means can be mainly attributed to true differences associated with herd, period or herd-period effects and only to a small degree be attributed to the vagaries of sampling. It was assumed in this study that the precision of the compositing sampling method most commonly used (two composites of seven or eight shipments in each period) was acceptable to the industry and that interest in a random sampling scheme was motivated by a desire to reduce the cost involved in sampling every shipment and in building and storing composite samples. Therefore the criterion of precision in this study was that a random sampling scheme should estimate herd-period means with a standard error equal to or less than the standard error of a composite estimate.

The sample size in a sampling scheme in which three traits are measured and the same precision is required for each trait is determined by the most variable trait. However, the criterion of precision used in this study could be different for each milk constituent and therefore the constituent which is now estimated most precisely could determine the sampling scheme if the scheme is to satisfy the criterion for all three milk constituents. However, milk fat is the only milk constituent currently used in establishing milk prices; therefore the precision of percent milk fat estimates was the only estimate whose precision could be assumed to have been accepted by the industry. For this reason the main emphasis in this study was on the precision of estimates of percent milk fat and a sampling scheme was deemed to be adequate if the criterion was met for this milk constituent. Protein or solids-not-fat content may be included in future milk pricing formulae in which case estimates of sampling variability would be useful to the industry, therefore, percent protein and lactose, the two main components of solids-not-fat, were included in this study.

The design of a sampling scheme to meet the specified criterion requires the estimation of the variance of current composite estimates and the within herd-period variance of milk constituent percent. The within herd-period variance (σ_w^2) measures the variability of estimates of the milk constituent percentages of each shipment of milk from a herd for a given period and can be written, after O'Keeffe [15]:

$$\sigma_w^2 = \sigma_d^2 + \sigma_s^2 + \sigma_t^2 \tag{1}$$

where

σ_d^2 biological variance--measures the variability due to true differences among shipments in milk constituent percentages;

σ_s^2 sampling variance--measures the variability among milk samples taken from the same shipment;

σ_t^2 testing variance--measures the variability among results of analyses (done at different times on different analysers) on the same sample.

Sampling and testing variances are due to procedures of estimation and may be combined:

$$\sigma_a^2 = \sigma_s^2 + \sigma_t^2 \tag{2}$$

where σ_a^2 is the within shipment variance--due to both sampling and testing.

Biological variance can be attributed mainly to:

1. day to day variability in both quantity and composition of milk produced by individual cows--this factor would give rise to random shipment to shipment fluctuations;
2. removals from or additions to the milking herd;

3. changes in routines and/or personnel associated with milking and handling the herd--these changes may occur regularly and give rise to cyclic fluctuations;
4. consistent directional trends across time in the percentage of a milk constituent, which may be due to the influence of changing seasons--this factor would be expected to give rise to a positive correlation between the percent of any milk constituent in one shipment with the percent of the same constituent in another shipment close to it in time; their correlation being a function of their distance apart and diminishing as the distance increases (serial correlations). If this factor is relatively important the biological variance would be expected to be lower in a short period than in a long period. In which case division of periods into sub-periods or strata and randomly selecting shipments for sampling from each strata (stratified random sampling) would be expected to reduce the standard error of the estimated herd-period mean as compared to simple random sampling.

Each milk shipment in a period is sampled to form a composite; therefore the variance of the estimated mean, between composite variability, is due to within shipment variability and to variability introduced by the procedures associated with the formation of a composite (compositing variance). Biological variance--which measures the variability among true shipment values--is not a component of the variance of composite estimates of a herd-period mean. The effects of the procedures associated with the formation of a composite sample could lead to consistent over- or under-estimation (bias) of herd-period means by composite samples.

Statistical Methods

Estimation of within herd-period variance. The fresh sample data of Experiment I were used to estimate the within herd-period variances of percent milk fat, protein and lactose using the analysis of variance. The linear mathematical model assumed was:

$$y_{ijk} = \mu + h_i + p_{j(i)} + w_{k(ij)} \qquad (3)$$

where

y_{ijk} the observed milk consistuent percent of the k^{th} shipment in the j^{th} period of the i^{th} herd;

μ the general mean;

h_i the effect associated with the i^{th} herd, $N(0,\sigma_h^2)$, σ_h^2 is the variance among herd means;

$p_{j(i)}$ the effect of the j^{th} period in the i^{th} herd, $N(0,\sigma_p^2)$, σ_p^2 is the variance among period means within herds;

$w_{k(ij)}$ the effect of the k^{th} shipment within the j^{th} period and i^{th} herd, $N(0,\sigma_w^2)$, σ_w^2 is the within herd-period variance.

The within strata variances for two, three and four strata per period were estimated from Experiment I for percent milk fat, protein and lactose using the analysis of variance. This analysis partitioned the within herd-period variance into; the variance among strata means and the residual within strata variance. If strata were a significant source of variation then stratified random sampling would be expected to be worthwhile. The linear mathematical model assumed was:

$$y_{ijkl} = \mu + h_i + p_{j(i)} + st_{k(ij)} + ws_{l(ijk)} \qquad (4)$$

where

y_{ijkl} the observed milk constituent percent of the l^{th} shipment in the k^{th} strata in the j^{th} period of the i^{th} herd;

$st_{k(ij)}$ the effect of the k^{th} strata in the j^{th} period of the i^{th} herd, $N(0,\sigma_{st}^2)$, σ_{st}^2 is the variance among strata means within periods and herds (σ_{st}^2 - two strata, $\sigma_{st'}^2$ - three strata and $\sigma_{st''}^2$ - four strata);

$ws_{l(ijk)}$ the effect of the l^{th} shipment in the k^{th} strata, j^{th} period and i^{th} herd, $N(0,\sigma^2_{ws})$, σ^2_{ws} is the within strata variance (σ^2_{ws} - two strata, $\sigma^2_{ws'}$ - three strata and $\sigma^2_{ws''}$ - four strata);

and the remaining symbols have been defined in equation 3.

Estimation of the components of within herd-period variance. The replicated sampling and testing data collected for one period in Experiment II was used to estimate the components, given in equation 1, of the within herd-period variance. The linear mathematical model assumed was:

$$y_{ijkl} = \mu + h_i + d_{j(i)} + s_{k(ij)} + t_{l(ijk)} \quad (5)$$

where

y_{ijkl} the observed milk constituent percent of the l^{th} test on the k^{th} sample from the j^{th} shipment of the i^{th} herd;

μ the general mean;

h_i the effect of the i^{th} herd, $N(0,\sigma^2_h)$;

$d_{j(i)}$ the effect of the j^{th} shipment of the i^{th} herd, $N(0,\sigma^2_d)$, σ^2_d is the biological variance;

$s_{k(ij)}$ the effect of the k^{th} sample from the j^{th} shipment of the i^{th} herd, $N(0,\sigma^2_s)$, σ^2_s is the sampling variance;

$t_{l(ijk)}$ the effect of the l^{th} test on the k^{th} sample from the j^{th} shipment of the i^{th} herd, $N(0,\sigma_t^2)$, σ_t^2 is the testing variance.

Mean squares were set equal to their expectations and the resulting equations were solved to obtain estimates of the components of the within herd-period variance.

The within strata biological variances for two, three and four strata per period were estimated from Experiment II for percent milk fat, protein and lactose. The linear mathematical model assumed was:

$$y_{ijklm} = \mu + h_i + st_{j(i)} + ds_{k(ij)} + t_{m(ijkl)} \qquad (6)$$

where

y_{ijklm} the observed milk constituent percent of the m^{th} test on the l^{th} sample from the k^{th} shipment in the j^{th} strata and the i^{th} herd;

$st_{j(i)}$ the effect of the j^{th} strata of the i^{th} herd, $N(0,\sigma_{st}^2)$, σ_{st}^2 is the variance among strata means within herds (σ_{st}^2 - two strata, $\sigma_{st'}^2$ - three strata and $\sigma_{st''}^2$ - four strata);

$ds_{k(ij)}$ the effect of the k^{th} shipment in the j^{th} strata and i^{th} herd, $N(0,\sigma_{ds}^2)$, σ_{ds}^2 is the within strata biological variance (σ^2_{ds} - two strata, $\sigma^2_{ds'}$ - three strata and $\sigma^2_{st''}$ - four strata);

and the remaining symbols have been defined in equation 5.

The difference between estimates of within herd-period variances (both with and without strata) obtained from Experiments I and II were tested by a two-tailed F-test.

Estimation of compositing variances. The data of Experiment III (triplicate composites and duplicate tests) were used to estimate compositing variances for percent milk fat, protein and lactose using the analysis of variance. The linear mathematical model assumed was:

$$y_{ijkl} = \mu + h_i + g_{j(i)} + c_{k(ij)} + t_{l(ijk)} \qquad (7)$$

where

y_{ijkl} is the observed milk constituent percent of the l^{th} test on the k^{th} composite in the j^{th} compositing period of the i^{th} herd;

μ the general mean;

h_i the effect of the i^{th} herd, $N(0,\sigma^2_h)$;

$g_{j(i)}$ the effect of the j^{th} compositing period in the i^{th} herd, $N(0,\sigma^2_g)$, σ^2_g is the variance among compositing period means within herds;

$c_{k(ij)}$ the effect of the k^{th} composite in the j^{th} compositing period of the i^{th} herd,

$N(0,\sigma_c^2)$, σ_c^2 is the compositing variance;

$t_{l(ijk)}$ the effect of the l^{th} test on the k^{th} composite in the j^{th} compositing period and the i^{th} herd, $N(0,\sigma_t^2)$, σ_t^2 is the testing variance.

The number of degrees of freedom associated with estimates of compositing variances from Experiment III were relatively small. Also composites were formed by the staff of the BCDA, Dairy Branch Laboratory; usually composites are formed in the laboratories of the milk plants to which the herd milk is shipped (as was the case in Experiment I). For these reasons estimates of compositing variances of percent milk fat, protein and lactose were obtained from Experiment I by an indirect method using estimates of sampling and testing variances from Experiment II. The linear mathematical model assumed was:

$$y_{ij} = \mu + h_i + p'_j + r_{ij} \qquad (8)$$

where

y_{ij} the mean milk constituent percent of the i^{th} herd for the j^{th} period, periods in this analysis were seven consecutive shipments (two weeks);

h_i the effect of the i^{th} herd, $N(0,\sigma_h^2)$;

p_j' the effect of the j^{th} seven shipment period, $N(0,\sigma_{p'}^2)$, $\sigma_{p'}^2$ is the variance among period means;

r_{ij} the joint effect of the i^{th} herd and the j^{th} period which includes the interaction between the i^{th} herd and the j^{th} period (hp_{ij}) and the random error (ei_{jk}).

Model 8 does not yield direct estimates of the variance of composite formation nor of the variance of a composite estimate. To estimate these variances three estimates of the mean percent composition of seven shipments of milk were used as dependent variables in model (8). These were: (a) the mean of seven fresh samples weighted by the weight of milk in each shipment; (b) the two-week composite estimate; (c) the mean of two one-week composites, weighted by the amount of milk represented by each composite. The difference between the residual variation of the fresh sample mean and the residuals of the two kinds of composites were equated to their expectations in order to solve for the desired estimates. The expectations of the within herd-period variances of the three estimates of the mean percent composition for a two week period can be written:

$$\sigma^2_{xf} = \frac{1}{n.}\sigma^2_s + \frac{1}{n.}\sigma^2_t \quad (9)$$

$$\sigma^2_{xc} = \sigma^2_c + \frac{1}{n.}\sigma^2_s + \sigma^2_t \quad (10)$$

$$\sigma^2_{x2c} = \sigma^2_{2c} + \frac{n_1+n_2}{n.^2}\sigma^2_s + \frac{n_1^2+n_2^2}{n.^2}\sigma^2_t \quad (11)$$

where

σ^2_{xf} the within herd-period variance of the mean of seven shipments each sampled and tested once;

σ^2_c the within herd-period variance of a two-week composite;

σ^2_{x2c} the within herd-period variance of the weighted mean of two one-week composites;

σ^2_c and σ^2_{2c} the variances associated with the formation of a two-week composite and two one-week composites respectively;

n_1 and n_2 the number of shipments in each of the two one-week composites ($n_1 = 3$ and $n_2 = 4$);

n. the number of shipment in a two-week composite. ($n. = n_1 + n_2 = 7$);

σ^2_s and σ^2_t defined in equation 1.

Equations 9, 10 and 11 represent the expectations of the random error of the residual mean square arising from r_{ij} in model 8; therefore, the expectations of the residual mean square for each kind of sampling may be written as follows:

$$\sigma^2_{rf} = \sigma^2_{ph(f)} + \frac{1}{n.}\sigma^2_s + \frac{1}{n.}\sigma^2_t \qquad (12)$$

$$\sigma^2_{rc} = \sigma^2_{ph(c)} + \sigma^2_c + \frac{1}{n.}\sigma^2_s + \sigma^2_t \qquad (13)$$

$$\sigma^2_{r2c} = \sigma^2_{ph(2c)} + \sigma^2_{2c} + \frac{n_1+n_2}{n.^2}\,\sigma^2_s + \frac{n_1^2+n_2^2}{n.^2}\,\sigma^2_t \qquad (14)$$

where

σ^2_{rf} is the residual mean square of the mean of seven shipments;

σ^2_{rc} is the residual mean square of two-week composites;

σ^2_{r2c} is the residual mean square of the mean of two one-week composites;

$\sigma^2_{ph(f)}$ is the variance associated with the herd-period interaction effect for fresh sampling;

$\sigma^2_{ph(c)}$ is the variance associated with the herd-period interaction effect for two-week composite sampling;

$\sigma^2_{ph(2c)}$ is the variance associated with the herd-period interaction effect for two one-week composite sampling;

σ^2_s and σ^2_t defined in equation 1.

The remaining symbols and the coefficients associated with variances have been defined in equations 9, 10 and 11.

If the interaction variance is assumed to be equal for all three kinds of sampling then the following equations hold:

For two-week composites:

$$\sigma^2_{rc} - \sigma^2_{rf} = \sigma^2_c + \frac{6}{7}\sigma^2_t \tag{15}$$

by rearrangement and substitution in equation 10;

$$\sigma^2_{xc} = \sigma^2_{rc} - \sigma^2_{rf} + \frac{1}{7}\sigma^2_s + \frac{1}{7}\sigma^2_t \tag{16}$$

For two one-week composites:

$$\sigma^2_{r2c} - \sigma^2_{rf} = \sigma^2_{2c} + \frac{18}{49}\sigma^2_t \tag{17}$$

and it follows that:

$$\sigma^2_{x2c} = \sigma^2_{r2c} - \sigma^2_{rf} + \frac{1}{7}\sigma^2_s + \frac{1}{7}\sigma^2_t \tag{18}$$

The estimates of sampling (σ^2_s) and testing (σ^2_t) variances obtained from the analysis of the data of Experiment II were used to solve equations 15 to 18 for the variance associated with the formation of composites (σ_c^2 or σ_{2c}^2) and for the variance of a composite estimate (σ_{xc}^2 or σ_{x2c}^2) of the period mean.

Serial correlations. The serial correlations, r_u, of y_{ij} with y_{ij+u} were calculated on a within herd basis; where y_{ij} is the observed milk constituent percent of the j^{th} shipment of the i^{th} herd and u varies from 1 to 14. The serial correlation coefficients were plotted on a correlogram versus u. The relationship between the serial correlations and u was estimated by fitting a fifth degree orthogonal polynomial, after Snedecor and Cochran [18]. The mathematical model assumed was:

$$r_u = b_o + \sum_{i=1}^{5} b_i u^i \qquad (19)$$

where

- r_u is the within herd serial correlation coefficient;
- b_o the population mean when u equals zero;
- u is the number of shipments separating the two shipments for which r_u was calculated ($u = 1,14$);
- b_i is the regression coefficient of r_u on u^i.

The graph of the equation which included only those powers in u which produced a significant reduction in the sums of squares was plotted on the correlogram.

Other statistical methods. Paired t-tests were used on the data of Experiment I to test for bias in composite estimates of herd-period (seven shipment periods) mean milk constituent percentage; the estimates obtained from each of the two kinds of composites were compared with the corresponding estimates obtained from fresh tests of milk samples taken from each shipment (fresh sample estimates). Fresh sample estimates were assumed to be the best unbiased estimates.

The level of significance was 0.05 for all statistical tests.

Standard errors of estimates of components of variance were calculated by the method of Anderson and Bancroft [1]. Standard errors of linear combinations of variances were computed after Welch [20].

1. RESULTS AND DISCUSSION

Estimates of Within Herd-period Variance and Components

The analysis of variance tables, showing the expectations of mean squares, of hierarchal models 3 and 5 used for the analysis of Experiments I and II are presented in Tables 1 and 2 respectively. The estimates from Experiment II of biological, sampling and testing variances for all three milk constituents are reported in Table 3. The within herd-period variances estimated from Experiment I are shown in column 5 of Table 3.

Sampling variance. The estimate of the sampling variance for percent milk fat was 0.00094 ± .00027 which was 7.1 percent of the total within herd-period variance. The estimates of sampling variance for percent protein and lactose were small and negative. These results indicated that drawing a milk sample, by the method used in the current study, was not an important source of variation for any of the three milk constituents. These findings agree with those of Dimick and Atherton [5] and Liska *et al*. [13] who found that sampling variance was low when bulk milk was properly agitated prior to taking a sample.

TABLE 1

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGES OF HERD BULK MILKS EXPERIMENT I

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	25	1127.81	45.11253	99.09*	$\sigma_w^2+k_2\sigma_p^2+k_3\sigma_h^2$
		180.47	7.21873	38.85*	
		37.70	1.50787	9.50*	
Periods (p)/h	289	131.58	0.45528	33.20*	$\sigma_w^2+k_1\sigma_p^2$
		53.70	.18581	23.60*	
		45.88	.15877	29.00*	
Shipments/p&h	4188	57.43	.01371		σ_w^2
		32.98	.00787		
		22.93	.00548		
Total	4502				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

k_1 = 14.29 k_2 = 14.33 k_3 172.88

TABLE 2

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGES OF HERD BULK MILKS EXPERIMENT II

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	21	158.53	7.54901	171.00*	$\sigma_t^2+k_1\sigma_s^2+k_3\sigma_d^2+k_4\sigma_h^2$
		51.53	2.45382	113.80*	
		10.35	0.49296	50.06*	
Shipments (d)/h	300	13.24	0.04416	5.40*	$\sigma_t^2+k_1\sigma_s^2+k_2\sigma_d^2$
		6.47	.02156	15.15*	
		2.95	.00985	2.95*	
Samples (s)/h&d	633	5.17	.00817	1.30*	$\sigma_t^3+k_1\sigma_s^2$
		0.90	.00142	.77	
		2.11	.00333	.84	
Tests/h, d&s	955	6.00	.00628		σ_t^2
		1.76	.00184		
		3.82	.00400		
Total	1909				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

$k_1 = 2$ $k_2 = 5.93$ $k_3 = 5.95$ $k_4 = 86.81$

TABLE 3

COMPONENTS OF WITHIN HERD-PERIOD VARIANCE (±S.E.) ESTIMATED FROM EXPERIMENT II
AND WITHIN HERD-PERIOD VARIANCE (±S.E.) ESTIMATED FROM EXPERIMENT I
PERIODS ARE FIFTEEN CONSECUTIVE SHIPMENTS

	Variance ($x10^{-2}$)				
Milk Constituent	Biological (1)	Sampling (2)	Testing (3)	Total[a] (4)	Total[b] (5)
% Milk fat	0.607±.061	0.094±.027	0.628±.029	1.329±.064	1.371±.030
% Protein	.340±.029	-.021±.006[c]	.167±.006	0.507±.031	0.787±.017
% Lactose	.110±.014	-.033±.013[c]	.373±.013	.483±.017	.548±.012

[a]total of columns one to three; i.e. $\sigma_w^2 = \sigma_d^2 + \sigma_s^2 + \sigma_t^2$.

[b]within herd-period variance (σ_w^2) estimated from experiment one.

[c]when estimates of sampling variance were negative the testing variance was estimated with increased degrees of freedom by combining the sums of squares for sampling and testing.

Testing variance. Estimates of testing variances and their standard errors for percent milk fat, protein and lactose were; 0.00628 ± .00029, 0.00167 ± .00006 and 0.00373 ± .00013 respectively (Table 3). Estimates of testing variance by Dunn [6] for the percent of the same three milk constituents (the analyses were performed in the same laboratory as the analyses in the current study) were; 0.00612, 0.00631 and 0.00505 respectively. In a review of automatic milk analysers, Green [10] reported that testing variances with IRMA under practical laboratory conditions were in the range 0.0036 to 0.0081 for all three milk constituents. Estimates of testing variance in the current study for percent milk fat and lactose (Table 3) fell in the range given by Green [10] and the estimate for percent milk fat closely agreed with the estimate by Dunn [6]. The estimate of testing variance for percent protein in the current study was smaller than the estimate by Dunn [6] and below the range reported by Green [10]. The difference between the estimate of percent protein testing variance by Dunn [6] and the estimate in the current study may indicate that testing variance varies from time to time under practical laboratory conditions. Testing variance as defined in the current study included; sample preparation, machine to machine variation and machine precision and thus represented the total variance associated with testing and sample handling procedures. The difference between estimates of testing variance in the current study and the variance between duplicates on IRMA of 0.0009 reported by Biggs [2] may be attributed to the contribution of the factors listed above other than machine precision.

The testing variances were 47.2, 32.9 and 77.2 percent (calculated from Table 3) of the total within herd-period variance for percent milk fat, protein and lactose respectively. Therefore testing was an important source of variation and consequently the number of determinations would have an important bearing on the variance of estimated period mean milk constituent percent. The variance of the mean percent lactose would depend mainly on the number of determinations and would be relatively independent of the sampling scheme.

Biological variance. Estimates of biological variances and their standard errors were 0.00607 ±.00061, 0.00340 ± .00029 and 0.00110 ± .00014 for percent milk fat, protein and lactose respectively. The biological variances were 45.7, 67.1 and 22.8 percent of the total within herd-period variance for percent milk fat, protein and lactose respectively.

Within herd-period variance. Estimates of the total within herd-period variance are shown in columns four (Experiment II) and five (Experiment I) of Table 3.

These estimates from the two experiments were not significantly different, by two-tailed F-tests, for percent milk fat or percent lactose. However, differences between the estimates were significant for percent protein. The components of within herd-period variance were defined to be; biological, sampling and testing variances (equation 1). Sampling variance was concluded to be very small (Table 3) for all three milk constituents. Therefore differences between the estimates of within herd-period variance obtained from Experiments I and II can mainly be attributed to differences in biological variance and/or in testing variance in the two experiments. If biological variances differ in the population, then random sampling schemes may have to be modified for different herds and/or different periods. Variability of testing variance would mean that the variance of estimates of herd-period mean milk constituent percentages cannot be accurately predicted for any sampling scheme.

Effects of Strata

Two shipments of milk which are close together in time can be expected to be more similar in milk constituent percent than two shipments which are more widely separated. (Materials and Methods).

Serial correlation. The sets of product moment correlations, r_u for pairs of shipments u shipments apart were calculated for values of u from one to fourteen on a within herd basis for twenty-three herds of Experiment I across thirteen periods for percent milk fat, protein and lactose. The results are presented in Table 4 and a correlogram shown in Figure 1.

The sets of product moment correlations were fitted to equation 19. The values of u^i (i = 1,5) in equation 19 were replaced by orthogonal polynomial coefficients from Fisher and Yates [8]. The reduction in sums of squares was tested as each successive term was added. As the objective was to find the polynomial of lowest degree that was a good fit, calculations were stopped when two successive additions were both non-significant (Tables 5A, 5B and 5C for percent milk fat, protein and lactose respectively). The coefficients in the resulting polynonomial equations were transformed to yield equations expressed in terms of u. These equations and graphs of these equations are shown in Figure 1.

Cochran [4] has shown that when a serial correlation exists in a population the standard error of the mean of a sample is reduced by using either stratified random or systematic sampling techniques. Cochran [4] also showed that when the correlogram is a straight line the variance of systematic sampling was equal to the variance of a stratified random sample, provided that there was no

TABLE 4

WITHIN HERD SERIAL CORRELATIONS[a] FOR PERCENT MILK FAT, PROTEIN AND LACTOSE

Number of Shipments Apart (u)	Serial Correlations			Number of Paired Values
	Milk Fat	Protein	Lactose	
1	0.826	0.736	0.749	4227
2	.777	.693	.724	4196
3	.742	.656	.684	4168
4	.707	.604	.648	4145
5	.685	.590	.633	4120
6	.663	.548	.631	4099
7	.632	.518	.597	4084
8	.603	.479	.590	4057
9	.572	.432	.564	4037
10	.560	.396	.531	4020
11	.541	.366	.528	3988
12	.519	.346	.505	3969
13	.502	.302	.465	3955
14	.478	.263	.438	3933

[a]the serial correlation, r_u of y_i with y_{i+u} computed on a within herd basis.

TABLE 5A

THE REDUCTION IN SUMS OF SQUARES DUE TO SUCCESSIVE TERMS IN THE POLYNOMIAL OF EQUATION 19. PERCENT MILK FAT SERIAL CORRELATIONS

Source	DF	SS	MS	F
Total	13	0.1513275		
Reduction to Linear	1	.1490156		
Deviations from Linear	12	.0023119	0.000193	773.5*
Reduction to Quadratic	1	.0019158		
Deviations from Quadratic	11	.0003961	.0000360	53.2*
Reduction to Cubic	1	.0001011		
Deviations from Cubic	10	.0002950	.0000295	3.4
Reduction to Quartic	1	.0000218		
Deviations from Quartic	9	.0002732	.0000304	0.7

*significant reduction of sums of squares.

TABLE 5B

THE REDUCTION IN SUMS OF SQUARES DUE TO SUCCESSIVE TERMS IN THE POLYNOMIAL OF EQUATION 19. PERCENT PROTEIN SERIAL CORRELATIONS

Source	DF	SS	MS	F
Total	13	0.2913929		
Reducation to Linear	1	.2906398		
Deviations from Linear	12	.0007531	0.0000628	4631.7*
Reduction to Quadratic	1	.0000859		
Deviations from Quadratic	11	.0006672	.0000607	1.4
Reduction to Cubic	1	.0000107		
Deviations from Cubic	10	.0006565	.0000657	0.2

*significant reduction in sums of squares.

TABLE 5C

THE REDUCTION IN SUMS OF SQUARES DUE TO SUCCESSIVE TERMS IN THE POLYNOMIAL OF EQUATION 19. PERCENT LACTOSE SERIAL CORRELATIONS

Source	DF	SS	MS	F
Total	13	0.1130053		
Reduction to Linear	1	.1112329		
Deviations from Linear	12	.0017724	0.0001477	753.1*
Reduction to Quadratic	1	.0000135		
Deviations from Quadratic	11	.0017589	.0001599	0.1
Reduction to Cubic	1	.0005543		
Deviations from Cubic	10	.0012046	.0001205	4.6

*significant reduction in sums of squares.

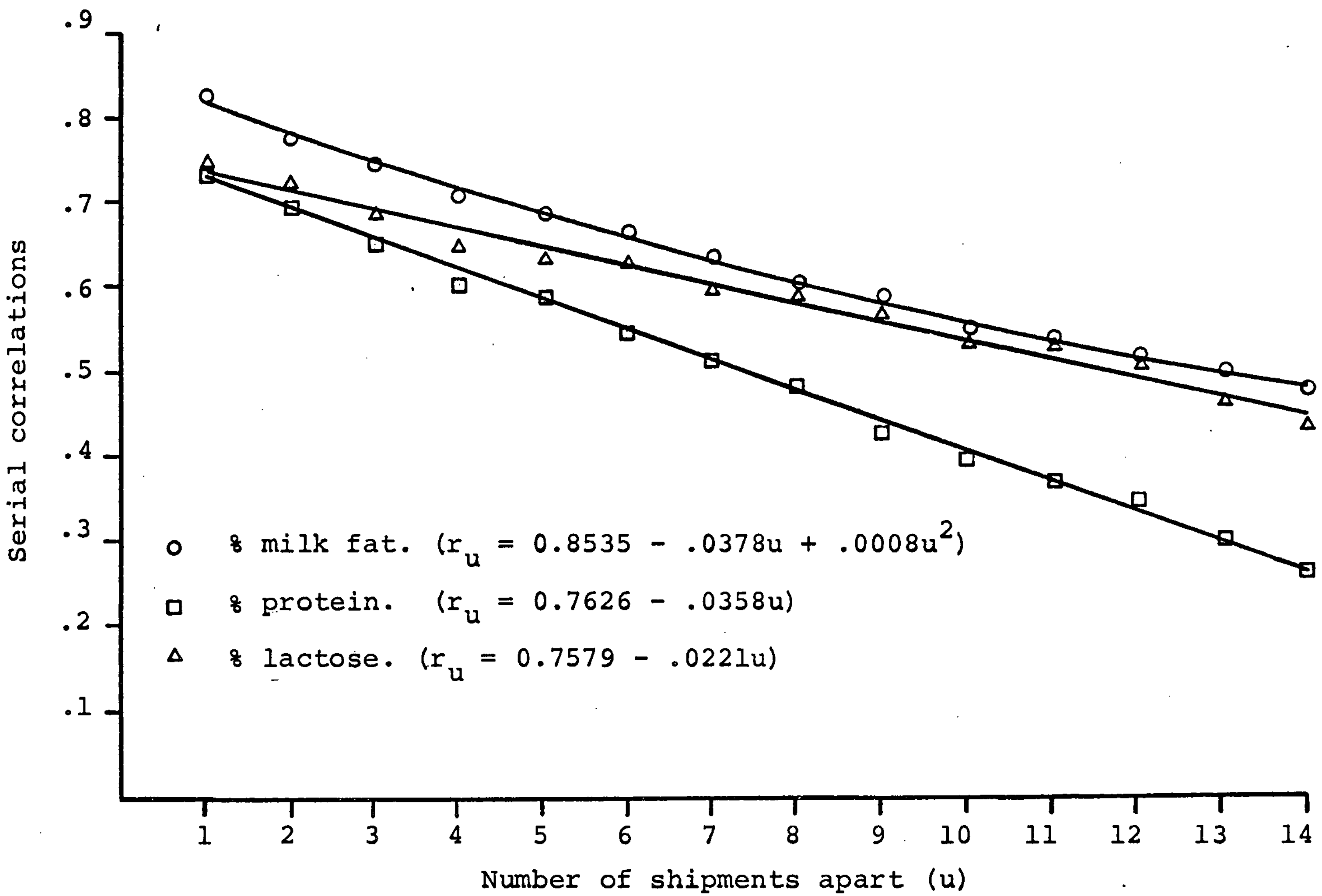

Figure 1 Serial correlations of percent milk fat, protein and lactose

periodic fluctuation in the population. However, when the correlogram is concave upward he reported that the variance of systematic sampling was less than the variance of stratified random sampling. When periodic variation exist in a population then the variance of systematic samples and the amount of bias in estimates provided by systematic samples depend on the relationship between the sampling frequency and the period of the fluctuations. Therefore, when fluctuations of unknown or variable period may exist in a population, stratified random sampling is to be preferred to systematic sampling. Cyclic fluctuations in milk constituent percentages may be present in the population currently under study (Material and Methods). The period of these fluctuations may differ between herds and also vary from time to time within a herd. Thus estimates of herd-period mean milk constituent percentages obtained by systematic sampling techniques could be biased; therefore, the use of systematic sampling was rejected in the current study.

The relationships between the serial correlations of percent protein and percent lactose and u were estimated as linear. The reduction in sums of squares due to the linear fit was 99.7 percent and 98.4 percent (calculated from Tables 5B and 5C respectively) of the total sums of squares of the serial correlations of percent protein and percent lactose respectively. The relationship between the serial correlations of percent milk fat and u contained a significant contribution due to the quadratic term in the equation. The reduction in sums of squares due fitting both linear and quadratic terms was 99.7 percent of the total sums of squares of the serial correlations of percent milk fat; the reduction due to fitting the linear term only was 98.4 percent of the total sums of squares (calculated from Table 5A). Therefore, although the graph describing the relationship between the serial correlations of percent milk fat and u was concave upwards, the departures from a linear relationship were relatively small.

These results indicated that, for all milk constituents, the serial correlations decreased regularly as u increased. Therefore, the variance of estimates of herd-period mean milk constituent percentages obtained by stratified random sampling would be expected to be smaller than the variance of estimates obtained by simple random sampling. The variance of the estimates would be expected to be lowest, for stratified random sampling schemes, when one observation is taken from each strata and when all strata are of equal size, Cochran [4].

Within Strata Variance

Estimates of the within strata variance were obtained from Experiment I using statistical model 4. Estimates of the within strata biological variance were obtained from Experiment II using statistical model 6. Three levels of stratification of fifteen shipment periods (one month) were used: (a) two strata, one of seven and one of eight shipments; (b) three strata of five shipments each; and (c) four strata with four shipments in three strata and three shipments in the fourth stratum. The analysis of variance table showing expectations of mean squares of Experiments I with two, three and four strata are presented in Tables 6A, 6B and 6C respectively. The results for Experiment II are presented in Tables 7A to 7C. Table 8 shows the biological variance for Experiment II and the within herd-period variance for both Experiments for all three milk constituents and for two, three and four strata. The effect of strata was a significant source of variation in all analyses. Therefore fitting strata reduced the magnitude of the within herd-period variance (within strata variance). The results were plotted in Figure 2 for percent milk fat, protein and lactose for both experiments. The values plotted in Figure 2 for no strata were from Table 3.

TABLE 6A

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGE OF HERD BULK MILKS EXPERIMENT I -- TWO STRATA PER PERIOD

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	25	1127.81	45.11253	99.09*	$\sigma^2_{ws}+k_4\sigma^2_{st}+k_5\sigma^2_p+k_6\sigma^2_h$
		180.47	7.21873	38.85*	
		37.70	1.50787	9.50*	
Periods (p)/h	289	131.58	0.45528	8.60*	$\sigma^2_{ws}+k_2\sigma^2_{st}+k_3\sigma^2_p$
		53.70	.18581	7.39*	
		45.88	.15877	10.05*	
Strata (st)/h & p	315	16.68	.05296	5.03*	$\sigma^2_{ws}+k_1\sigma^2_{st}$
		7.92	.02515	3.89*	
		4.98	.01580	3.41*	
Shipments/h,p & st	3873	40.75	.01052		σ^2_{ws}
		25.06	.00647		
		17.95	.00464		
Total	4502				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

$k_1 = 7.08$ $k_2 = 7.22$ $k_3 = 14.29$ $k_4 = 7.24$ $k_5 = 14.33$ $k_6 = 172.88$

TABLE 6B

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGE OF HERD BULK MILKS. EXPERIMENT I -- THREE STRATA PER PERIOD

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	25	1127.81	45.11253	99.09*	$\sigma^2_{ws'}+k_4\sigma^2_{st'}+k_5\sigma^2_p+k_6\sigma^2_h$
		180.47	7.21873	38.85*	
		37.70	1.50787	9.50*	
Periods (p)/h	289	131.58	0.45528	11.74*	$\sigma^2_{ws'}+k_2\sigma^2_{st'}+k_3\sigma^2_p$
		53.70	.18581	10.21*	
		45.88	.15877	14.02*	
Strata (st)/h & p	629	24.38	.03876	4.18*	$\sigma^2_{ws'}+k_1\sigma^2_{st'}$
		11.45	.01821	3.01*	
		7.21	.01132	2.55*	
Shipments/h,p & st	3559	33.04	.00928		$\sigma^2_{ws'}$
		21.53	.00605		
		15.81	.00444		
Total	4502				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

$k_1 = 4.75$ $k_2 = 4.81$ $k_3 = 14.29$ $k_4 = 4.82$ $k_5 = 14.33$ k_6 172.88

TABLE 6C

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGE OF HERD BULK MILKS. EXPERIMENT I -- FOUR STRATA PER PERIOD

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	25	1127.81	45.11253	99.09*	$\sigma^2_{ws''}+k_4\sigma^2_{st''}+k_5\sigma^2_p+k_6\sigma^2_h$
		180.47	7.21873	38.85*	
		37.70	1.50787	9.50*	
Periods (p)/h	289	131.58	0.45528	15.74*	$\sigma^2_{ws''}+k_2\sigma^2_{st''}+k_3\sigma^2_p$
		53.70	.18581	11.15*	
		45.88	.15877	16.37*	
Strata (st)/h & p	944	27.31	.02893	3.12*	$\sigma^2_{ws''}+k_1\sigma^2_{st''}$
		15.73	.01666	3.13*	
		9.15	.00970	2.28*	
Shipments/h,p & st	3244	30.12	.00928		$\sigma^2_{ws''}$
		17.25	.00532		
		13.78	.00425		
Total	4502				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

$k_1 = 3.55$ $k_2 = 3.66$ $k_3 = 14.29$ $k_4 = 3.66$ $k_5 = 14.33$ $k_6 = 172.88$

TABLE 7A

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGES OF HERD BULK MILKS
EXPERIMENT II -- TWO STRATA PER PERIOD

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	21	158.53 51.53 10.35	7.54901 2.45382 0.49296	46.30* 21.33* 16.82*	$\sigma_t^2+k_1\sigma_s^2+k_3\sigma_{ds}^2+k_5\sigma_{st}^2+k_6\sigma_h^2$
Strata (st)/h	22	3.59 2.53 0.64	.16304 .11505 .02931	4.69* 8.12* 3.53*	$\sigma_t^2+k_1\sigma_s^2+k_3\sigma_{ds}^2+k_4\sigma_{st}^2$
Shipments (ds)/h & st	278	9.66 3.94 2.31	.03474 .01416 .00831	4.25* 9.95* 2.49*	$\sigma_t^2+k_1\sigma_s^2+k_2\sigma_{ds}^2$
Samples (s)/h,st & ds	633	5.17 0.90 2.11	.00817 .00142 .00334	1.30* .77 .84	$\sigma_t^2+k_1\sigma_s^2$
Tests/h,st,ds & s	955	6.00 1.76 3.82	.00628 .00184 .00400		σ_t^2
Total	1909				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

$k_1 = 2$ $k_2 = 5.93$ $k_3 = 5.95$ $k_4 = 43.16$ $k_5 = 43.66$ $k_6 = 86.81$

TABLE 7B

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGES OF HERD BULK MILKS
EXPERIMENT II -- THREE STRATA PER PERIOD

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	21	158.53 51.53 10.35	7.54901 2.45382 0.49296	75.54* 25.17* 4.97*	$\sigma_t^2+k_1\sigma_s^2+k_3\sigma_{ds'}^2+k_5\sigma_{st'}^2+k_6\sigma_h^2$
Strata (st)/h	44	4.40 4.29 1.36	.09993 .09748 .03094	2.89* 11.45* 4.97*	$\sigma_t^2+k_1\sigma_s^2+k_3\sigma_{ds'}^2+k_4\sigma_{st'}^2$
Shipments (ds/h & st	256	8.85 2.18 1.59	.03456 .00851 .00622	4.23* 5.98* 1.86*	$\sigma_t^2+k_1\sigma_s^2+k_2\sigma_{ds'}^2$
Samples (s)/h,st & ds	633	5.17 0.90 2.11	.00817 .00142 .00334	1.30* 0.77 0.84	$\sigma_t^2+k_1\sigma_s^2$
Tests/h,st,ds & s	955	6.00 1.76 3.82	.00628 .00184 .00400		σ_t^2
Total	1909				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

$k_1 = 2$ $k_2 = 5.93$ $k_3 = 5.95$ $k_4 = 28.88$ $k_5 = 29.06$ $k_6 = 86.81$

TABLE 7C

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGES OF HERD BULK MILKS. EXPERIMENT II -- FOUR STRATA PER PERIOD

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	21	158.53	7.54901	80.68	$\sigma^2_t+k_1\sigma^2_s+k_3\sigma^2_{ds''}+k_5\sigma^2_{st''}+k_6\sigma^2_h$
		51.53	2.45382	38.43*	
		10.35	0.49296	19.02*	
Strata (st)/h	66	6.18	.09357	3.10*	$\sigma^2_t+k_1\sigma^2_s+k_3\sigma^2_{ds''}+k_4\sigma^2_{st''}$
		4.21	.06386	6.63*	
		1.71	.02592	4.88*	
Shipment (ds)/h & st	234	7.07	.03020	3.70*	$\sigma^2_t+k_1\sigma^2_s+k_2\sigma^2_{ds''}$
		2.25	.00963	6.77*	
		1.24	.00531	1.57*	
Samples (s)/h,st & ds	633	5.17	.00817	1.30*	$\sigma^2_t+k_1\sigma^2_s$
		0.90	.00142	0.77	
		2.11	.00334	0.84	
Tests/h,st,ds & s	955	6.00	.00628		σ^2_t
		1.76	.00184		
		3.82	.00400		
Total	1909				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

$k_1 = 2$ $k_2 = 5.93$ $k_3 = 5.95$ $k_4 = 21.50$ $k_5 = 22.31$ $k_6 = 86.81$

TABLE 8

WITHIN HERD-PERIOD TOTAL VARIANCE FROM EXPERIMENT I AND BIOLOGICAL AND TOTAL VARIANCE FROM EXPERIMENT II WITH NO STRATA AND TWO, THREE AND FOUR STRATA FOR PERCENT MILK FAT, PROTEIN AND LACTOSE

Number of Strata	(Components of Variance ($x10^{-2}$))		
	Experiment II		Experiment I
	Biological	Total	Total
A. Milk fat percent			
None	0.607±.061	1.329±.064	1.371±.030
Two	.448±.051	1.170±.054	1.052±.024
Three	.445±.052	1.167±.056	0.928±.022
Four	.372±.048	1.094±.049	.928±.023
B. Protein percent			
None	0.340±.029	0.507±.031	0.787±.017
Two	.215±.020	.382±.021	.647±.015
Three	.119±.013	.286±.014	.605±.014
Four	.139±.015	.306±.016	.532±.013
C. Lactose percent			
None	0.110±.014	0.483±.017	0.548±.012
Two	.084±.012	.457±.016	.464±.011
Three	.049±.010	.422±.014	.444±.011
Four	.033±.009	.406±.014	.425±.011

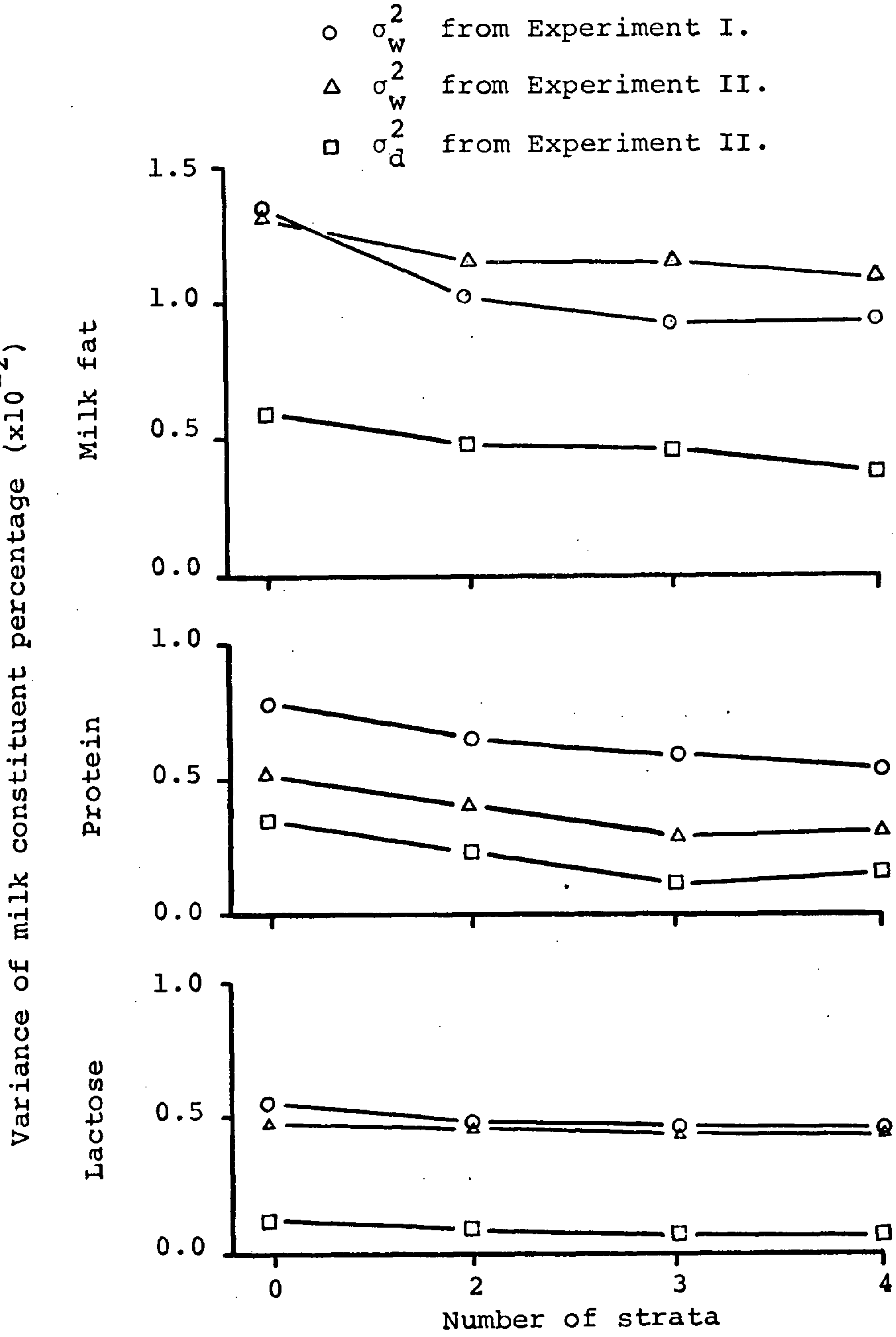

Figure 2 Within herd-period biological and total variance estimated with no strata and with two, three and four strata per period for percent milk fat protein and lactose

Within strata variance of percent milk fat. The estimates of within herd-period variance--no strata--of percent milk fat obtained from the two experiments were not significantly different by an F-test. However the estimates of within herd-period variance of percent milk fat calculated within strata (within strata variance) from Experiment I were significantly lower than the estimates from Experiment II. Stratification is expected to reduce the effect of time trends on the magnitude of the within herd-period variance (Materials and Methods). Differences between estimates of herd-period variance obtained from the two experiments can be attributed to either differences between biological variances or to differences between testing variances in the two sets of data. Stratification would be expected to reduce biological variance only: testing variance (within shipment variation) would not be altered by stratification. Therefore, the results indicated that time trends (averaged across thirteen periods) in Experiment I may have been a more important source of variation than time trends in the single period in Experiment II. Thus directional changes in milk constituent percentages may be greater in some periods (seasons) than in others. In which case, unless stratification can effectively stabilize within herd-period variances, it may be necessary to take more milk samples in some seasons than in other seasons if the same level of precision is to be achieved for all seasons.

Within strata variances of percent protein. The estimates of the within herd-period variances, with and without stratification, for percent protein obtained from Experiment I were all significantly lower than the corresponding estimates obtained from Experiment II (Table 8 and Figure 2) by an F-test. The difference between the estimates was relatively constant for all levels of stratification. The results indicated, by use of the same reasoning that was applied to the results for the within herd-period variance of milk fat percentage, that testing variance was different between the two data sets. If testing variance changes from time to time then predictions of the variance of herd-period mean milk constituent percentages cannot be made accurately. However, variability of testing variance would affect the precision of all milk sampling schemes; although, the magnitude of the change in precision may not be the same for all schemes. The difference between biological variance and within herd-period variance was assumed to be equal to testing (and sampling) variance, equation 1 . If biological variances were approximately the same in the two data sets, then testing variance in Experiment I could be approximated by the difference between biological variance estimated from Experiment II and within herd-period variance from Experiment I. The estimate of testing variance for percent protein in Experiment I obtained by the average of these differences was 0.00440 (calculated from Table 8).

Within strata variance of percent lactose. The differences between the estimates of within herd-period variances, with and without stratification, for percent lactose obtained from Experiments I and II were non-significant.

Variability of Estimates from Various Sampling Schemes

The variance of the mean ($\sigma^2_{\bar{x}}$) estimated by drawing a simple random sample of n shipments from a period of N shipments can be written:

$$\sigma^2_{\bar{x}} = \frac{1}{n} [\sigma^2_d (1-\frac{n}{N}) + \sigma^2_s + \sigma^2_t] \quad (20)$$

where the symbols have been defined in equation 1. The variance of the mean estimated by a stratified random sample can be written:

$$\sigma^2_{\bar{x}} = \sum_{i=1}^{m} W^2_i \frac{1}{n_i} [\sigma^2_{ds} (1-\frac{n_i}{N_i}) + \sigma^2_s + \sigma^2_t] \quad (21)$$

where

W_i the weight attached to the i^{th} strata and is equal to the number of shipments in the i^{th} strata divided by the total number of shipments in the period;

m the number of strata in the period;

n_i the number of observations drawn from the i^{th} strata;

N_i the number of shipments in the i^{th} strata;

σ^2_{ds} the within strata biological variance, which is assumed equal for all strata.

With equal strata size and equal number of observations per stratum, equation 21 reduces to:

$$\sigma^2_{\bar{x}} = \frac{1}{n'm} [\sigma^2_{ds}(1 - \frac{n'm}{N}) + \sigma^2_3 + \sigma^2_t) \qquad (22)$$

where

n' is the number of observations per stratum.

The other symbols remain as previously defined.

The estimates of the variances obtained from the analyses of Experiment II were substituted into the appropriate equations to calculate the variance of the mean and the 99% confidence interval about the mean for various sampling schemes (Table 9). The results showed that the reduction in the confidence limits by stratification was

TABLE 9

PREDICTED VARIANCE AND 99% CONFIDENCE INTERVAL OF THE MEAN OF FRESH SAMPLES OF VARYING SIZES DRAWN FROM A PERIOD OF 15 SHIPMENTS FOR PERCENT MILK FAT, PROTEIN AND LACTOSE SIMPLE AND STRATIFIED RANDOM SAMPLING

Number[a]	% Milk Fat		% Protein		% Lactose	
	$\sigma^{2b}_{\bar{x}}$	99%CL[c]	$\sigma^{2b}_{\bar{x}}$	99%CL[c]	$\sigma^{2b}_{\bar{x}}$	99%CL[c]
A. Simple random sampling						
1	1.289	±.293	0.484	±.180	0.475	±.178
2	0.624	±.203	.231	±.124	.234	±.125
3	.403	±.164	.146	±.099	.154	±.104
4	.292	±.139	.104	±.083	.113	±.087
5	.225	±.122	.079	±.072	.089	±.077
6	.181	±.111	.062	±.064	.073	±.070
7	.149	±.100	.050	±.058	.062	±.064
8	.126	±.092	.041	±.052	.053	±.059
9	.107	±.084	.034	±.047	.046	±.056
15	.048	±.057	.011	±.027	.025	±.041
B. Stratified random sampling						
			Two Strata			
2	0.558	±.193	0.178	±.109	0.224	±.122
4	.264	±.133	.082	±.074	.109	±.085
6	.182	±.110	.050	±.057	.071	±.069
8	.126	±.092	.034	±.047	.052	±.059
			Three Strata			
3	0.359	±.154	0.087	±.076	0.137	±.096
6	.165	±.105	.036	±.049	.067	±.067
9	.100	±.082	.024	±.040	.044	±.054
			Four Strata			
4	0.254	±.130	0.068	±.067	0.101	±.082
8	.114	±.087	.030	±.044	.049	±.057

[a]number of samples per period for both simple and stratified random sampling

[b]variance of the mean $\times 10^{-2}$.

[c]99% confidence interval of the mean.

relatively small and diminished as n increased. The reduction in the confidence intervals was greatest for percent protein and least for percent lactose. These results can be attributed mainly to two factors. Firstly, biological variance (between shipment variation) was the only component of the within herd-period variance that could be expected to be reduced by stratification; sampling and testing variances (within shipment variation) would not be altered. Therefore, stratification would be expected to reduce the confidence intervals to a greater extent for those milk constituents for which biological variance was a major component of the within herd-period variance. Secondly, the finite population correction factor applied only to the biological variance therefore the contribution of biological variance to the standard error of the mean would be reduced more rapidly as sample size increased than the contribution of sampling and testing variances. Thus for relatively large n the contribution of biological variance to the standard error would be small and therefore the effect of any reduction in the magnitude of biological variance by stratification on the standard error would diminish as n increased. Stratification could still be worthwhile if it resulted in a reduction in the frequency of large deviations from the true mean by eliminating the probability of drawing all observations from either the beginning or the end of a period. Although large deviations may occur with relatively low frequency their occurrance could be of concern to the individual milk producer as his payment for the period's milk shipments are based on the results of the estimate of the mean percent milk fat.

Composite Sampling

Variance of composites--Experiment III. The criterion of precision in the current study was that a random sampling scheme should estimate herd period means at least as precisely as two-week composite sampling. Experiment III was designed to provide estimates of the standard error of herd-period means estimated by the mean of two two-week composites; one of seven shipments and one of eight shipments.

Each shipment of milk was sampled in the formation of composites; therefore, the variance of a composite estimate was entirely attributable to procedures of estimation: (1) sampling, (2) testing and (3) formation of a composite sample. The data of Experiment III were analysed by statistical model 7 to obtain estimates of the variance associated with the formation of composite samples. The analyses of variance table showing the expectation of mean squares is presented in Table 10. Compositing was not a significant source of variation for any of the three

TABLE 10

ANALYSIS OF VARIANCE OF MILK CONSTITUENT PERCENTAGE OF HERD BULK MILKS
EXPERIMENT III: ESTIMATE OF COMPOSITING VARIANCE

Source	DF	SS[a]	MS[a]	F[a]	EMS
Herds (h)	20	20.7105 7.2451 1.3535	1.03553 0.36225 .06768	48.40* 22.08* 6.93*	$\sigma_t^2+2\sigma_c^2+6\sigma_g^2+12\sigma_h^2$
Periods (g)/h	21	0.4493 .3445 .2050	.02140 .01641 .00976	15.41* 3.08* 4.53*	$\sigma_t^2+2\sigma_c^2+6\sigma_g^2$
Composites (c)/h & g	84	.1167 .4474 .1812	.00139 .00533 .00216	1.14 0.91 .47	$\sigma_t^2+2\sigma_c^2$
Tests/h,g & c	126	.1530 .7343 .5808	.00121 .00583 .00461		σ_t^2
Total	251				

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

milk constituents studied. Estimates of compositing and testing variance (Table 11) for all three milk constituents were obtained by equating the mean squares to their expectations and solving the resulting equations.

The estimate of the variance associated with the formation of a composite sample was low (0.000087 ± .00013) for percent milk fat, and was low and negative (-.000251 ± .000546) for percent protein. The estimate for percent lactose while negative was relatively large in absolute value (-.001226 ± .000332).

Estimates of testing variance obtained from Experiment II (Table 3) were compared with estimates obtained from Experiment III (Table 11). F-tests showed that the estimate of testing variance for percent milk fat was significantly lower in Experiment III then in Experiment II; the estimate of percent protein testing variance was significantly higher. The estimates of testing variance for percent lactose were not significantly different between the two experiments. These results support the conclusions, based on the comparison of the estimates of within herd-period variances obtained from Experiments I and II, that the testing variance for percent protein may vary from time to time. The results from Experiment III indicated that testing variance for percent milk fat also may vary from time to time. The estimates of testing variances from Experiment II were based on analyses done

TABLE 11

ESTIMATES OF COMPOSITING AND TESTING VARIANCE
EXPERIMENT III

Milk Constituent	Variance $(x10^{-2})$	
	Compositing	Testing
% Milk fat	0.0087 ± .0130	0.1214 ± .0152
% Protein	-.0251 ± .0546	.5627 ± .0547
% Lactose	-.1226 ± .0332	.3629 ± .0352

over a period of one month. The analyses of the composite in Experiment III were done on two days (two weeks apart) in the same month as the analyses for Experiment II. Therefore, testing variances would appear to be subject to considerable short-term fluctuations.

Variance of composites -- Experiment I. The results from the analyses of Experiment III indicated that the formation of composites is not an important source of variation of composite sample estimates of the period mean percent composition. However, as the number of degrees of freedom associated with these estimates was relatively low the data from Experiment I were analysed by statistical model 8. The residuals from these analyses were equated to their expectations to yield estimates of the variance of composite formation and the variance of a composite estimate as shown in equations 9 to 18 (Statistical Methods).

The analysis of variance tables (model 8) for fresh sample, two-week composite and two one-week composite estimates are presented in Tables 12A, 12B and 12C respectively. The estimates of sampling and testing variance (Table 3) and the residual mean squares (Tables 12A, 12B, and 12C) were used to solve equations 15 through 18 for the variance of composite estimates and the variance

TABLE 12A

ANALYSIS OF VARIANCE OF MILK CONSTITUENT PERCENTAGES FITTING HERDS AND PERIODS (MODEL 8) EXPERIMENT I FRESH SAMPLE EXTIMATES

Source	DF	SS[a]	MS[a]	F[a]
Herds	25	151.938 23.982 4.955	6.07751 0.95930 .19821	393.3* 110.1* 55.3*
Periods	25	11.579 4.158 5.241	.46316 .16632 .20962	30.0* 19.1* 58.5*
Residual	564	8.714 4.915 2.023	.01545 .00871 .00359	
Total	614			

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively

*significant source of variation.

TABLE 12B

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGES FITTING HERDS AND PERIODS (MODEL 8) EXPERIMENT I TWO-WEEK COMPOSITE ESTIMATES

Source	DF	SS[a]	MS[a]	F[a]
Herds	25	139.792	5.59169	256.9*
		23.808	0.95233	89.1*
		4.304	.17216	24.7*
Periods	25	11.190	.44760	20.6*
		5.958	.23831	22.3*
		4.623	.18492	26.5*
Residual	564	12.275	0.02176	
		6.024	0.01068	
		3.931	.00697	
Total	614			

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

TABLE 12C

ANALYSES OF VARIANCE OF MILK CONSTITUENT PERCENTAGES FITTING HERDS AND PERIODS (MODEL 8) EXPERIMENT I TWO ONE-WEEK COMPOSITE ESTIMATES

Source	DF	SS[a]	MS[a]	F[a]
Herds	25	143.551 25.356 5.138	5.74205 0.97424 0.20552	311.6* 96.4* 39.6*
Periods	25	9.601 4.791 4.513	0.38404 0.19164 0.18052	20.8* 20.0* 34.8*
Residual	564	10.394 5.700 2.924	0.01843 0.01011 0.00518	
Total	614			

[a]the three values listed for each source of variation are for percent milk fat, protein and lactose respectively.

*significant source of variation.

associated with the formation of a composite sample for both types of composite samples used in Experiment I. The estimates of compositing variances for all three milk constituents (Table 13) were smaller, or only slightly larger, than their standard errors for both two-week composites and two one-week composites. These results support the conclusions, based on Experiment III, that the formation of a composite is not an important source of variation. If sampling and compositing variances are both small then testing was the most important source of variation of composite estimates (equation 10).

Calculation of the Criterion of Precision

The estimates of the variance of a composite for a two-week compositing period (Table 13) were used to calculate the variance of the mean of two two-week composites ($\sigma^2_{\bar{x}c}$). The equation can be written:

$$\sigma^2_{\bar{x}c} = \sum_{i=1}^{m} \frac{n_i^2}{N^2} \sigma^2_{xc_i} \qquad (23)$$

where

m is the number of composite samples in a period (month);

n_i is the number of shipments in the i^{th} compositing period;

TABLE 13

VARIANCE OF COMPOSITES ($x10^{-2}$)

	Two-week		Two One-week	
Milk Constituent	σ^{2}_{xc}[a]	σ^{2}_{c}[b]	σ^{2}_{x2c}[c]	σ^{2}_{2c}[d]
% Milk fat	0.735±.151	0.093±.195	0.401±.137	0.067±.140
% Protein	.221±.078	.054±.081	.164±.076	.079±.077
% Lactose	.410±.048	.018±.041	.212±.035	.022±.036

[a]variance of a seven shipment composite (for a two week period).

[b]variance associated with the formation of a two week composite.

[c]variance of the mean (for a two week period) of two composites of three or four shipments each.

[d]variance associated with the formation of two one-week composites.

N is the number of shipments in a period (month);

$\sigma^2_{xc_i}$ variance of the i[th] composite, (which is defined in equation 10 for two-week composites).

The variance of the mean of two two-week composite are presented in Table 14. These values were the maximum variances of the means of random samples allowable if the criterion of precision was to be met.

Composite Sampling versus Random Sampling

Accuracy of composites. Experiment I was used to test the accuracy of composite sampling. Three estimates of herd mean milk constituent percentages for each two-week period were obtained from Experiment I. The best unbiased estimate of each herd two-week mean was considered to be the mean of the fresh samples, weighted by the amount of milk in each shipment (fresh sample estimates). Paired t-tests were used to test differences between fresh sample estimates and means estimated by; (a) the observed value of a two-week composite and (b) the mean of two one-week composites weighted by the amount of milk represented by each composite.

The results (Table 15) indicated that percent milk fat was significantly underestimated by both types of composites; the difference between fresh and both types

TABLE 14

VARIANCES OF HERD-PERIOD MEAN MILK CONSTITUENT PERCENT ESTIMATED BY TWO TWO-WEEK COMPOSITES PER PERIOD

Milk Constituent	Variance ($x10^{-2}$) of Estimates of Herd-Period Means
% Milk fat	0.3675
% Protein	0.1105
% Lactose	0.2050

TABLE 15

PAIRED t-TEST OF DIFFERENCES BETWEEN THE FRESH[a] ESTIMATE OF A TWO WEEK PERIOD MEAN AND BOTH KINDS OF COMPOSITE ESTIMATES

Milk Constituent	Fresh vs. Two-week			Fresh vs. Two One Week		
	Diff.[b]	S.D.	t	Diff.[b]	S.D.	t
% Milk fat	-.045	0.113	-10.6*	-.045	0.080	-14.5*
% Protein	0.023	.085	7.0*	0.003	.060	1.3
% Lactose	.010	.064	4.2*	.010	.057	4.3*
DF	694			656		

[a]mean of fresh samples from seven consecutive shipments.

[b]mean difference: composites minus fresh.

*significant difference.

of composite estimates was 0.045 percent milk fat. Percent protein was overestimated by two-week composites (the difference was 0.025 percent protein), but two one-week composite estimates were not significantly different from fresh estimates. Both types of composites overestimated percent lactose by 0.010 percent lactose. Herrmann and Anderson [11] and Preston [17] also reported that percent milk fat was lower in composite samples than fresh samples.

Estimation of sample size. The criterion used in the current study was that the standard error of an estimate from a random sample should be at least as low as the standard error of the estimate from composite samples currently in use. The number of random samples (n) required to give a predetermined variance of the mean can be found by rearranging equation 20 to yield:

$$n = \sigma_w^2 / (\sigma_{\bar{x}}^2 + \sigma_{\frac{d}{N}}^2) \qquad (24)$$

where

$\sigma_{\bar{x}}^2$ is the predetermined variance of the mean;

N is the number of shipments in the period;

and the remaining symbols have been defined in equation 1. This equation also holds for stratified random sampling if the strata size are equal, the sampling fraction is the same for all strata and σ_d^2 is defined as the within strata biological variance.

The appropriate predetermined variances ($\sigma_{\bar{x}}^2$) in the current study for each milk constituent were the variances of the mean milk constituent percentages, for a period of fifteen (N) shipments, estimated by two two-week composite samples (Table 14). The numbers of random samples required per month for each milk constituent were calculated from equation 24 by using these variances as $\sigma_{\bar{x}}^2$ and the estimates of biological variances (σ_d^2) and within herd-period variances (σ_w^2) from Experiment II (Table 3). The results of the calculations are presented in Table 16. The criterion specified that random sampling should be at least as precise as composite sampling; therefore, as the number of samples has to be a whole number, the values in Table 16 should be increased to the next whole number. This calculation showed that four simple random samples per period would be predicted to estimate herd-period mean milk constituent percentages with a variance less than the variance of current estimates which are based on two two-week composites per period.

The variance of composite estimates is due entirely to procedures of estimation; i.e. compositing, sampling and testing. The proceeding analyses indicate that testing variance is the most important (Tables 3 and 11) of the three. The variance of estimates based on random sampling are due to both procedures of estimation (sampling and testing) and to biological or day to day differences of true shipment means. Therefore the number of random samples required to give an estimate of the mean with a precision equal to that of composite samples depends on the relationship between biological variance and procedural variance. An expression defining this relationship can be derived by equating the expectations of the variance of a composite estimate to the expectations of the variance of the mean of a random sample of n shipments.

The variance of the mean of m composites collected over a period of N shipments with each composite representing the same number of shipments (N/m) can be written:

$$\sigma^2_{\bar{x}c} = \frac{1}{m}(\sigma^2_c + \sigma^2_t) + \frac{\sigma^2_s}{N} \qquad (25)$$

Equating this equation to equation 20 and rearranging yields:

$$n = m(r + 1)/\{1 + \frac{1}{\sigma^2_a}[\sigma^2_c + \sigma^2_s(\frac{m}{N} - 1)] + \frac{mr}{N}\} \qquad (26)$$

where

n is the number of random samples required to give an estimate of the mean with precision equal to that from a composite scheme;

N is the number of shipments in the period for which an estimate of the mean is desired;

m is the number of compositing periods in the current composite scheme;

σ_c^2 is the variance associated with the formation of a composite;

r is the ratio σ_d^2 / σ_a^2 ;

and the remaining symbols have been defined in equations 1 and 2.

The term:

$$\frac{1}{\sigma_a^2} [\sigma_c^2 + \sigma_s^2(\frac{m}{N} - 1)] \qquad (27)$$

in the denominator of equation 26 reduces to zero if:

$$\sigma_c^2 = (1 - \frac{m}{N}) \sigma_s^2 \qquad (28)$$

and is near zero if both σ_c^2 and σ_s^2 are small relative to σ_t^2, that is, if the main procedural source of variation is testing. If the term shown in formula 27 can be assumed to be very close to or equal to zero then equation 26 reduces to:

$$n = m(r + 1) / (1 + \frac{mr}{N}) \qquad (29)$$

TABLE 16

ESTIMATES OF SAMPLE SIZE REQUIRED IF THE VARIANCE OF THE MEAN IS TO EQUAL THE VARIANCE OF THE MEAN OF TWO TWO-WEEK COMPOSITES

Method[a]	Percent Milk fat	Percent Protein	Percent Lactose
A	3.26	3.81	2.27
B	3.31	4.78	2.49

[a]Method.

[A]Calculated from formula 24, with σ_w^2 and σ_d^2 estimated from experiment two.

[B]calculated from formula 29.

If σ_c^2 is large the use of the simplified equation will tend to overestimate n. If σ_s^2 is large then n will be underestimated. The advantage of using equation 29 rather than equation 24 is that only the ratio of biological to testing plus sampling variance need to be known or estimated in order to calculate the number of random samples needed to replace a compositing scheme with a random sampling scheme of equal precision. Estimates of sample size calculated by equation 29 are presented in Table 16 and agree well with those calculated by equation 24.

The number of samples required (n in formula 29) were graphed (Figure 3) versus the ratio of biological to testing variance (r in equation 29) for two, three and four composites per period. The graph can be used to find the number of samples required (n) for various values of r for three compositing schemes.

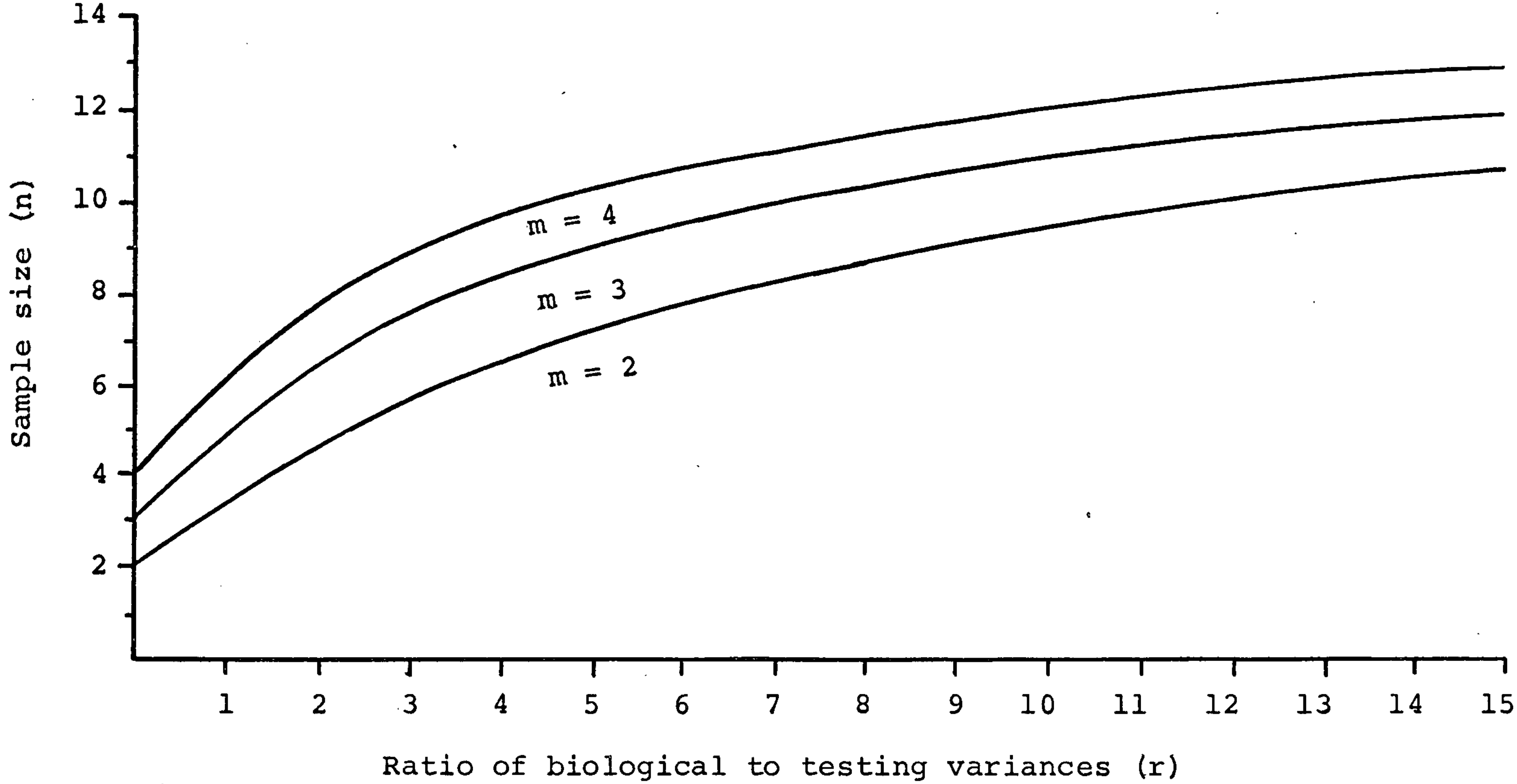

Figure 3 The number of samples required (n) for random sampling to equal the precision of composite sampling for various ratios of biological to testing variance (r) calculated from equation 29

1. CONCLUSIONS

Estimates of sampling variance for all three milk constituents were very small relative to the total within herd-period variance of milk constituent percentages. From these results it can be concluded that the method of sampling bulk milk used in this study introduced little variation into estimates of milk constituent percentages of bulk milk.

Estimates of compositing variances were also small for all three milk constituents. As both compositing and sampling variances were concluded to be small, then it follows that testing is the main source of variation in composite sample estimates of herd-period mean milk constituent percentages.

Estimates of testing variances for percent milk fat and percent protein obtained from Experiment II were significantly different from the corresponding estimates obtained from Experiment III. Therefore it can be concluded that testing variances vary from time to time in the laboratory. If this conclusion is true then statistically valid predictions of the variances of estimates, obtained from any sampling scheme, of herd-period mean milk constituent percentages cannot be made. However, practical considerations demand that reasonable limits be placed on the magnitudes of testing variances so that the variability of estimates obtained from various sampling schemes can be at least approximated. More data would be required to estimate the amount of variation in testing variances.

The analyses indicated that biological variances may vary from time to time or from herd to herd. Variation in biological variances would mean that a random sampling scheme may have to be modified for different seasons or herds. However, it may be possible to associate differences in biological variances with seasons or with variables associated with herds (e.g. quantity of milk shipped) and thereby simplify the modification of random sampling schemes to suit different herds or seasons.

Two-week composite samples were concluded to yield biased estimates of all three milk constituent percentages. One-week composites were biased estimates of percent milk fat and lactose. Random sampling would be expected to yield unbiased estimates. Therefore, it was concluded that estimates of herd-period mean milk constituent percentages obtained from sampling four randomly selected shipments would be at least as precise as, and more accurate (unbiased) than, estimates obtained from two two-week composites. It was also concluded that stratified random sampling (with one sample per strata) would reduce the variability of these estimates. As both testing and biological variances may vary, these conclusion apply to the average condition and may not be valid for all herds or periods.

PART 2

2. INTRODUCTION

Estimates of population variances associated with bulk milk sampling and testing were obtained in Part 1. These estimates were used to predict the variability of estimates of herd-period mean milk constituent percentages under various sampling schemes. Part 2 of this thesis investigated some of the practical problems associated with random sampling schemes.

The results presented in Part 1 indicated that the variance of estimates of herd-period mean milk constituent percentages obtained by random sampling would be expected to be no greater than the variance of estimates obtained by composite sampling if four random milk samples were taken each period. However, the variances of estimates obtained by each of these two sampling schemes were attributed to different sources. The variance of estimates obtained by composite sampling was attributed to procedures of estimation (sampling, testing and compositing). The variance of estimates obtained by random sampling was attributed to true differences among shipments (biological variance) and to procedures of estimation (sampling and testing).

The magnitude of variances associated with procedures of estimation may vary from time to time (Part 1) but would be expected to be essentially the same for all herds at a given time. The magnitude of biological variance, however, is not necessarily the same for all herds. Therefore, the variance of estimates obtained by composite sampling would be similar for all herds; but the variance of estimates obtained from random sampling could differ among herds. Thus, although a particular random sampling scheme may, on the average, meet a specified acceptable level of precision, estimates of herd-period mean milk constituent percentages obtained by this scheme could be much more variable for some herds than others. Each estimate is economically important to the individual producer; therefore, ideally, the variance of estimates should be the same for all herds. If the variability of estimates obtained by random sampling is to be approximately equal for all herds then different sampling schemes may be necessary for some herds.

For the above reasons the data of Experiment I were used to estimate if herds differed in within herd-period variances of milk constituent percentages and if these differences (if any) were large enough to warrant different sampling schemes for certain herds.
The data of Experiment 2 were also used to estimate if herd-period variances can be predicted

from easily measured variables associated with herds.

Within herd-period variances of milk constituent percentages may differ among seasons. These differences may be attributed to changes in either biological variance or testing variance (Part 1). If biological variance is higher at certain seasons, sampling frequency should be increased in these seasons. On this basis sampling schemes may need to be modified not only for certain herds but also for certain periods.

Mistakes in sample identification, analyses etc., can occasionally be made; therefore, test results should be systematically checked for gross errors. Factors involved in checking the results from a random sampling scheme are discussed in Part 2.

2. MATERIALS AND METHODS

Source of Data

The fresh sample data collected in Experiment I (defined in Part 1) from the twenty-three herds that shipped milk throughout the thirteen periods of Experiment I were used for the analyses in Part 2. All periods in Part 2 were of fifteen consecutive milk shipments (approximately one month).

Statistical Methods

The average amount of milk in each shipment in each of the thirteen periods was calculated (the arithmetic average of all shipments). The mean milk constituent percent for each period was also calculated (the average of fresh samples each weighted by the amount of milk it represented).

The within herd-period variances (σ_w^2) of percent milk fat, protein and lactose were calculated for each herd-period subclass for twenty-three herds and thirteen periods. The within herd-period variances of percent milk fat, protein and lactose were also calculated with two, three and four strata per period by dividing the sums of squares pooled over the strata by the pooled degrees of freedom. Frequency distributions of these variances were constructed. Herd and period mean within herd-period variances of percent milk fat, protein and lactose were calculated for each herd and each period by dividing the pooled sums of squares (pooled over periods for herd means and pooled over herds for period means) by the pooled degrees of freedom.

Regression analyses. Simple and multiple regression techniques were used to estimate the effects of herd size (measured by average milk shipment weight) and milk constituent percentages on the within herd-period variance of all three milk constituents. The sampling distribution of estimates of variances are not expected to be normal; a logarithmic transformation is expected to yield a normal distribution, Snedecor and Cochran [18]. Therefore the distributions of the within herd-period variances and the natural logarithm ($\log_e$) of these variances were both tested for skewness and kurtosis by the method of Snedecor and Cochran [18]. The regressions were fitted overall, within herds and within periods.

Simple linear regressions of the $\log_e$ of the within herd-period variances of percent milk fat, protein and lactose were fitted on each of four independent variables which are defined as follows:

M_{ij} the mean weight (kg.) of milk in each shipment of the i^{th} herd and j^{th} period;

F_{ij} the mean percent milk fat associated with the i^{th} herd and j^{th} period;

P_{ij} the mean percent protein associated with the i^{th} herd j^{th} period;

L_{ij} the mean percent lactose associated with the i^{th} herd and j^{th} period.

The overall simple linear regression model assumed was:

$$y_{ij} = b_0 + b_1 X_{ij} + e_{ij} \quad (30)$$

where

y_{ij} the natural logarithm of the within herd-period variance of each milk constituent percent of the i^{th} herd and j^{th} period;

b_0 the population mean when X_{ij} equals zero;

b_1 the simple regression coefficient of y_{ij} on X_{ij};

X_{ij} was set equal to M_{ij}, F_{ij}, P_{ij} and L_{ij} in turn;

e_{ij} the random error, $N(0,\sigma_e^2)$.

The within herd simple linear regression model assumed was:

$$y_{ij} = b_0 + h_i + b_1X_{ij} + e_{ij} \quad (31)$$

where

b_0 the population mean when equal frequencies exist in all subclasses and when X_{ij} equals zero;

h_i the effect associated with the i^{th} herd;

b_1 the within subclass simple regression coefficient of y_{ij} on X_{ij};

e_{ij} the random error $N(0,\sigma_e^2)$;

and the remaining symbols were defined in equation 30.

The within period simple linear regression model assumed was:

$$y_{ij} = b_0 + p_j + b_1X_{ij} + e_{ij} \quad (32)$$

where

p_j the effect associated with the j^{th} period;

e_{ij} the random error, $N(0,\sigma_e^2)$;

y_{ij} defined in equation 30;

and the remaining symbols were defined in equation 31.

Multiple regressions of the $\log_e$ of within herd period variances of percent milk fat, protein and lactose were fitted on the four independent variables.

The overall multiple regression model assumed was:

$$y_{ij} = b_0 + b_1M_{ij} + b_2F_{ij} + b_3P_{ij} + b_4L_{ij} + e_{ij} \qquad (33)$$

where

b_0 the population mean when M_{ij}, F_{ij}, P_{ij} and L_{ij} all equal zero;

b_1 the partial regression coefficient of y_{ij} on M_{ij};

b_2 the partial regression coefficient of y_{ij} on F_{ij};

b_3 the partial regression coefficient of y_{ij} on P_{ij};

b_4 the partial regression coefficient of y_{ij} on L_{ij};

e_{ij} the random error $N(0,\sigma_e^2)$;

y_{ij} was defined in equation 30 and M_{ij}, F_{ij}, P_{ij} and L_{ij} were defined on page 87.

The within herd multiple regression model assumed was:

$$y_{ij} = b_0 + h_i + b_1M_{ij} + b_2F_{ij} + b_3P_{ij} + b_4L_{ij} + e_{ij} \qquad (34)$$

where

b_0 the population mean when equal frequencies exist in all subclasses and when M_{ij}, F_{ij},

P_{ij} and L_{ij} all equal zero;

h_i was the effect associated with the i^{th} herd;

b_1 was the within subclass partial regression coefficient of y_{ij} on M_{ij};

b_2 was the within subclass partial regression coefficient of y_{ij} on F_{ij};

b_3 was the within subclass partial regression coefficient of y_{ij} on P_{ij};

b_4 was the within subclass partial regression coefficient of y_{ij} on L_{ij};

e_{ij} was the random error, $N(0,\sigma_e^2)$;

y_{ij} was defined in equation 30;

and M_{ij}, F_{ij}, P_{ij} and L_{ij} were defined on page 87.

The within period multiple regression model assumed was:

$$y_{ij} = b_0 + p_j + b_1M_{ij} + b_2F_{ij} + b_3P_{ij} + b_4L_{ij} + e_{ij} \quad (35)$$

where

p_i was the effect associated with the j^{th} period;

e_{ij} was the random error, $N(0,\sigma_e^2)$;

y_{ij} was defined in equation 30;

and the remaining symbols were defined in equation 34.

Simple and partial regression coefficients were tested for significance by t-tests.

Differences among adjusted herd and period means were tested by F-tests of the reduction in the residual sums of squares of the overall (both simple and multiple) regressions obtained by fitting the within subclass regression models. The F-value is calculated, after Freese [9] as follows:

$$F_{s-1,\nu} = \frac{SSE - SSE'}{s - 1} / MSE' \qquad (36)$$

where

- SSE the residual sums of squares from the overall regression models, (equation 30 for simple and equation 33 for multiple regressions);
- SSE' the residual sums of squares from the within subclass regression models (equations 31 and 32 for simple and equations 34 and 35 for multiple regressions);
- MSE' residual mean square from within subclass regression models;
- s the number of subclasses in the within subclass regression models;
- ν the number of degrees of freedom associated with the error mean squares in the within subclass regression models;

The above F-test is identical to the F-test of the main effects in the analysis of covariance in the one-way classification.

Within period regressions measured the extent to which herd differences in the within herd-period variance of milk constituent percent can be attributed to herd differences in the independent variables. The within herd regressions measured the extent to which changes in the value of the independent variables in a herd were associated with changes in the within herd-period variance of milk constituent percent.

All possible samples for seven sampling schemes were computer generated from the data of Experiment I. Frequency distributions of the absolute value of the deviation of each sample from the fresh mean were constructed for percent milk fat and percent protein. The fresh mean was the mean of all fresh samples, weighted by the amount of milk in the shipment, in a period. The seven sampling schemes were for one to four random milk samples per period drawn without stratification and with stratification for those schemes with more than one milk sample per period. The schemes were:

1. One shipment sampled per period.
2. Two shipments sampled per period:
   (a) No strata
   (b) Two strata (one of seven and one of eight shipments).
3. Three shipments sampled per period:

(a) No strata

(b) Three strata (of five shipments each).

4. Four shipments sampled per period:

(a) No strata

(b) Four strata (three of four shipments and one of three shipments).

2. RESULTS AND DISCUSSION

Period Effects on Milk Shipment Weight and Milk Constituent Percentages

Figure 4 shows the mean milk shipment weight and mean milk constituent percent for each of the thirteen periods used in this study.

Percent milk fat. Percent milk fat dropped in the spring, remained at a relatively low level through the summer and climbed slowly to its peak value in mid-winter.

Percent protein. Percent protein increased in the spring, dropped off in the late summer, climbed to a peak in the autumn and then dropped slowly to a stable winter level.

Percent lactose. Percent lactose, which was less variable than either percent milk fat or protein, was lowest in the summer and autumn.

Milk shipment weight. The amount of milk shipped per herd was highest in the spring and early summer and dropped to its lowest levels in late summer and autumn.

The effect of season on the composition and level of production of herd milk can mainly be attributed to two factors. Firstly, to the stage of lactation of the

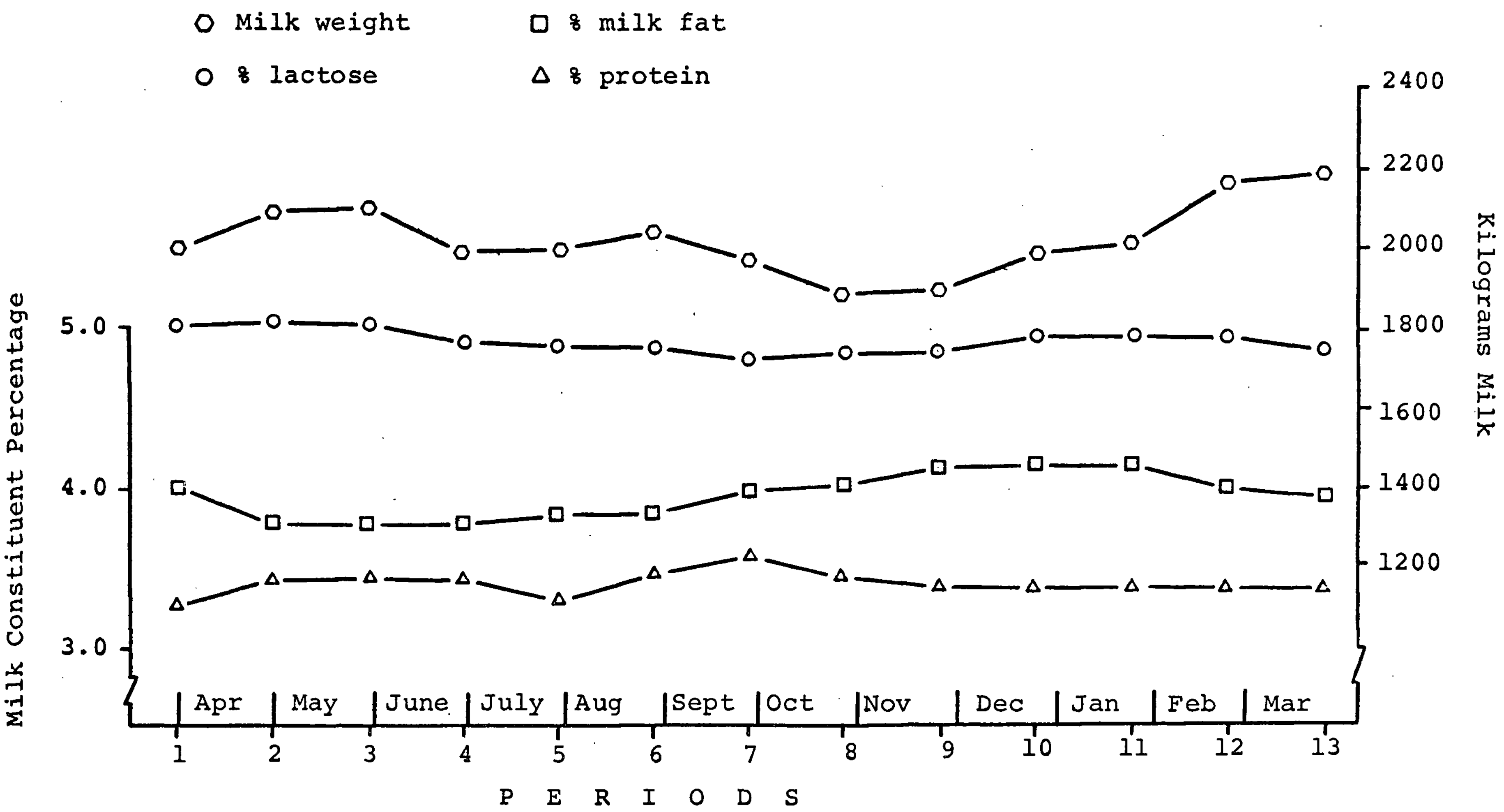

Figure 4 Period average milk constituent percentages and milk shipment weight for thirteen periods

cows in a herd in a particular season (i.e. the calving distribution) and secondly to the effect of season of the year on milk production and composition on cows at all stages of lactation. These factors can fluctuate from year to year and therefore the seasonal effects may vary. However, the seasonal trends reported in the current study agree with those reported by Waite and Robertson [19], Johnson et al. [12] and Boswell et al. [3].

Transformations

Table 17 shows the results of the tests for skewness and kurtosis in the distributions of the transformed and untransformed within herd-period variances of percent milk fat, protein and lactose. In all cases the untransformed data showed significant skewness and kurtosis, however, after transformation both skewness and kurtosis were non-significant.

Regression Analyses

The $\log_e$ transformed within herd-period variances of percent milk fat, protein and lactose, calculated with no strata and with four strata per period, were fitted as dependent variables to regression models 30 to 35. All results are presented in the transformed scale, so that regression coefficients measure the change in the

TABLE 17

TESTS OF NORMALITY OF THE DISTRIBUTION OF WITHIN HERD-PERIOD VARIANCES BEFORE AND AFTER LOGARITHMIC TRANSFORMATION

Milk Constituent	Skewness			Kurtosis		
	Untransformed	Transformed	Stan.Dev.	Untransformed	Transformed	Stan.Dev.
% Milk fat	4.85*	0.20	0.139	40.91*	0.05	0.277
% Protein	2.01	.12	.139	5.10*	.10	.277
% Lactose	2.37*	.03	.139	8.61*	.01	.277

*significant skewness or kurtosis.

$\log_e$ of within herd-period variance of a given milk constituent percentage associated with a unit change in an independent variable.

Within Herd-Period Variance of Percent Milk Fat

The estimates of the regression coefficients, t-tests of the coefficients and the proportion of the sums of squares (R^2) accounted for by the regression equations are shown in Table 18A for simple linear regressions and multiple linear regression; overall, within period and within herd for the $\log_e$ of the within herd-period variance of percent milk fat. F-tests of the differences among herds and among periods are also shown in Table 18A. The results for the $\log_e$ of the within herd-period variance of percent milk fat with four strata per period are shown in Table 18B.

Milk shipment weight. The overall simple linear regression of the $\log_e$ of the within herd-period percent milk fat variance on the average weight (kg.) of milk shipped was significant and the regression coefficient was $(-.264 \pm .083)\times 10^{-4}$; the within period regression was also significant $(-.249 \pm 0.76)\times 10^{-4}$ but the within herd regression was non-significant $(-.295 \pm .318)\times 10^{-4}$. These results indicated that herds shipping larger amounts

TABLE 18A

SIMPLE (SLR) AND MULTIPLE LINEAR (MLR) REGRESSION COEFFICIENTS FOR THE REGRESSION OF THE LOGARITHM OF THE WITHIN HERD-PERIOD VARIANCE OF PERCENT MILK FAT ON KILOGRAMS OF MILK, PERCENT MILK FAT, PROTEIN AND LACTOSE - NO STRATA

	Overall		Within Period			Within Herd		
	b±S.E.	R^2	b±S.E.	R^2 [c]	F	b±S.E.	R^2 [c]	F
SLR[a]								
Milk wt.[b]	-.264±.083*	3.29	-.249±.076*	3.65	6.28*	-.295±.318	0.32	2.14*
% Fat	0.128±.082	0.83	0.197±.076*	2.31	6.62*	-.753±.238*	3.56	3.01*
% Protein	.724±.186*	4.91	.792±.177*	6.67	6.73*	0.372±.380	0.98	1.91*
% Lactose	.889±.318*	2.58	.529±.378	0.69	5.61*	.897±.406*	1.76	2.46*
MLR[a]								
Milk wt.[b]	-.249±.086*	2.54	-.189±.080*	1.81		-.578±.327	1.09	
% Fat	-.422±.127*	3.37	-.377±.147*	2.13		-.658±.240*	2.62	
% Protein	1.184±.282*	5.36	1.331±.327*	5.34		0.694±.385	1.13	
% Lactose	1.040±.325*	3.13	0.405±.414	0.31		1.101±.444*	2.14	
MLR equation		11.43		10.09	5.67*		6.56	1.79*

[a]degrees of freedom: SLR; 294,282 and 272; MLR; 291, 279 and 269 for overall, within periods and within herds respectively.

[b]regression coefficients x10^{-4}.

*significant: regression coefficients by t-tests and differences among levels (within subclass models) by F-tests.

[c]R^2 calculated on the total within subclass sums of squares

TABLE 18B

SIMPLE AND MULTIPLE LINEAR REGRESSION COEFFICIENTS PERCENT MILK FAT WITH FOUR STRATA PER PERIOD

	Overall		Within Period			Within Herd		
	b±S.E.	R^2	b±S.E.	$R^{2\,c}$	F	b±S.E.	$R^{2\,c}$	F
SLR[a]								
Milk wt.[b]	-.308±.080*	4.75	-.288±.074*	5.12	5.72*	-.610±.299*	1.51	2.89*
% Fat	0.282±.078*	4.25	0.317±.073*	6.24	6.22*	-.154±.229	0.17	2.77*
% Protein	.774±.179*	5.97	.878±.171*	8.55	6.43*	-.116±.359	.04	2.48*
% Lactose	.636±.311*	1.41	.547±.370	0.77	5.42*	0.339±.387	.28	3.24*
MLR[a]								
Milk wt.[b]	-.238±.084*	2.47	-.190±.078*	1.90		-.756±.315*	2.10	
% Fat	-.035±.124	0.02	-.025±.143	0.01		-.102±.232	0.07	
% Protein	0.615±.278*	1.53	0.776±.320*	1.89		0.040±.372	.01	
% Lactose	.501±.319	0.77	.040±.405	0.01		0.609±.428	.74	
MLR equation		9.00		10.49	6.03*		2.48	2.32*

[a]degrees of freedom: SLR; 294, 282 and 272; MLR; 291, 279 and 269 for overall, within periods and within herds respectively.

[b]regression coefficients x10^{-4}.

*significant: regression coefficients by t-tests and differences among levels (within subclass models) by F-tests.

[c]R^2 calculated on the total within subclass sums of squares

of milk were associated with low within herd-period variance of milk fat percent. But, increased milk shipments by a particular herd were not associated with a significant reduction in the within herd-period variance of milk fat percent. The range of milk shipment weights was much greater (therefore the standard error of the regression coefficient was much smaller) for both the overall and the within period regressions than for the within herd regression.

Milk fat percent. The overall simple linear regression of the $\log_e$ of the within herd-period variance of milk fat percent on the average milk fat percent was non-significant; the regression coefficient was 0.128±.082. Both the within period and the within herd regressions were significant; the regression coefficients were 0.197±.076 and -.753±.238 respectively. These results indicated that high percent fat herds tend to have large variances of percent milk fat; but that within herds, periods of low percent milk fat (spring, see Figure 4) were associated with high variance of milk fat percent. The increase in the within herd-period variance of milk fat percent that was associated with periods of low milk percent may be due to the relatively rapid decline of milk fat percent associated with the advent of spring grazing. A consistent directional change in a milk constituent percentage across time would be expected to increase within herd-period variance of the milk constituent percentage.

Protein percent. The overall and the within period simple linear regressions were significant and the regression coefficients were 0.724 ± .186 and 0.792 ± .177 respectively. However, the within herd regression was non-significant; the coefficient was 0.372 ± .380. Thus herds with high percent protein had higher than average within herd-period variance of milk fat percent but changes in protein content within a herd were not significantly associated with changes in the variance of milk fat percent.

Lactose percent. The simple linear regression coefficients were significant for overall (0.889 ± .318) and within herds (0.897 ± .406) regression equations, but the within period regression coefficient (0.529 ± .378) was non-significant. These results indicated that increases in percent lactose within a herd were associated with an increase in the within herd-period variance of milk fat percentage, but that differences between herds in percent lactose were not significantly associated with differences in the variance of milk fat percent.

The F-tests of the differences among levels were significant for both within subclass regression models and for all independent variables used. These results indicated that differences among both herd and period means were significant when the independent variables were held constant (i.e. differences exist among herd means even after adjustment for the effects of herd size).

Overall multiple linear regression. All coefficients differed significantly from zero by a t-test when all four independent variables were included in the overall multiple linear regression model. The coefficients for average milk weight and milk fat percent were positive while those for percent protein and lactose were negative (Table 18A). The model accounted for 11.43 percent of the sums of squares of the dependent variable.

Within period multiple linear regression. Three of the independent variables were significant; average milk shipment weight, percent milk fat and percent protein when all four independent variables were included in the within period multiple linear regression model. These three independent variables were also significant when fitted singly in the simple linear regression model. However, the sign of the coefficient for percent milk fat changed from positive in simple linear regression to negative when the remaining three independent variables were held constant. The within period multiple regression model accounted for 10.09 percent of the total within period sums of squares of the dependent variable. Differences among periods in the dependent variable were significant by the F-test when the independent variables were held constant (Table 18A).

Within herd multiple linear regression. Two of the independent variables, percent milk fat and percent lactose, were significant sources of variation when the within herd multiple linear regression model was fitted. These two independent variables were also the only significant sources of variation when fitted in the simple linear regression models. The within herd multiple regression model accounted for 6.56 percent of the total within herd sums of squares of the dependent variable. Differences among herds in the dependent variable were significant by the F-test when the independent variables were held constant.

Within strata variance of milk fat percent. The $\log_e$ of the within herd-period variances of percent milk fat, calculated on a pooled within four strata basis, were fitted as dependent variables to the same regression models.

The regression coefficients estimated when the variance was calculated without stratification (Table 18A) were not significantly different from the regression coefficients estimated with four strata per period (Table 18B). However, for the independent variables of percent milk fat and percent lactose the within herd regression coefficients (both multiple and simple) were not significant when the variance was calculated with stratification but the regression coefficients were significantly different from zero when the variance was calculated without stratification. For milk shipment weight the within herd regression coefficients were not significant when the variance was calculated without stratification but were significant when the variance was computed with four strata per period.

Differences among herd and period means, tested by the F-test of the difference in levels of the within herd and within period regressions, were significant when the variance was calculated with four strata per period.

Within herd-period variance of a milk constituent percentage can be mainly attributed to two factors (Materials and Methods); (1) random day-to-day variations in the milk constituent percent and (2) directional changes in the milk constituent percent across time. The second factor (time trends) would be expected to account for more of the within herd-period variation in long periods than in short periods (strata). The magnitude of the random component would not be expected to change with length of periods. The results of the regression analyses (strata vs. no strata) indicated that the relationships between the within herd-period variance of milk fat percent and the independent variables can be mainly attributed to the magnitude of the random part of the within herd-period variance of milk fat percent.

Within Herd-Period Variance of Percent Protein

The $\log_e$ of the within herd-period variance of percent protein, calculated with four strata per period and without stratification, were used as dependent variables in regression models 30 to 35.

The estimates of the regression coefficients, t-tests of the coefficients and the proportion of the sums of squares (R^2) accounted for by the regression equations are shown in Table 19A for simple linear regressions and multiple linear regression; overall, within period and within herd for the $\log_e$ of the within herd-period variance of percent protein. F-tests of the differences among herds and periods are also shown in Table 19A. The results for the $\log_e$ of the within herd-period variance of percent protein calculated with four strata per period are shown in Table 19B.

Milk shipment weight. Milk shipment weight was a significant source of variation for the overall and the within period simple linear regression equations. The regression coefficients were $(-.306 \pm .071)\text{x}10^{-4}$ and $(-.293 \pm .064)\text{x}10^{-4}$ respectively. Milk shipment weight was not a significant source of variation for the within herd regression. These results indicated that herds

TABLE 19A

SIMPLE (SLR) AND MULTIPLE LINEAR (MLR) REGRESSION COEFFICIENTS FOR THE REGRESSION OF THE LOGARITHM OF THE WITHIN HERD-PERIOD VARIANCE OF PERCENT PROTEIN ON KILOGRAMS MILK PERCENT MILK FAT, PROTEIN AND LACTOSE - NO STRATA

	Overall		Within Period			Within Herd		
	b±S.E.	R^2	b±S.E.	$R^{2\,c}$	F	b±S.E.	$R^{2\,c}$	F
SLR[a]								
Milk wt.[b]	-.306±.071*	5.92	-.293±.060*	7.07	7.60*	-.524±.278	1.30	1.50
% Fat	0.131±.070	1.16	0.171±.065*	2.41	7.61*	-.296±.211	0.72	2.11*
% Protein	0.572±.161*	4.11	0.512±.152*	3.87	7.14*	0.697±.330*	1.62	2.05*
% Lactose	-.362±.278	0.57	-.075±.322	0.02	7.05*	-.636±.357	1.16	2.26*
MLR[a]								
Milk wt.[b]	-.257±.075*	3.64	-.244±.068*	4.18		-.440±.289	0.83	
% Fat	-.153±.111	0.60	-.051±.125	0.05		-.374±.213	1.10	
% Protein	0.710±.247*	2.58	0.479±.280	0.96		0.620±.342	1.18	
% Lactose	-.383±.284	0.57	-.304±.354	0.24		-.410±.394	0.39	
MLR equation		9.33		8.86	6.99*		4.18	1.38

[a]degrees of freedom: SLR; 294, 282 and 272; MLR; 291, 279 and 269 for overall, within periods and within herds respectively.

[b]regression coefficients $x10^{-4}$.

*significant: regression coefficients by t-tests and differences among levels (within subclass models) by F-tests.

[c]R^2 calculated on the total within subclass sums of squares

TABLE 19B

SIMPLE AND MULTIPLE LINEAR REGRESSION COEFFICIENTS
PERCENT PROTEIN WITH FOUR STRATA PER PERIOD

	Overall		Within Period			Within Herd		
	b±S.E.	R^2	b±S.E.	$R^{2\,c}$	F	b±S.E.	$R^{2\,c}$	F
SLR[a]								
Milk wt.[b]	-.188±.075*	2.06	-.167±.067*	2.16	7.99*	-.729±.292*	2.25	1.74*
% Fat	0.212±.073*	2.81	0.194±.066*	2.93	7.99*	0.121±.224	0.11	1.33
% Protein	0.477±.169*	2.63	0.566±.156*	4.48	8.56*	-.065±.352	0.01	1.34
% Lactose	0.057±.290	0.01	0.058±.331	0.01	7.95*	-.116±.380	0.03	1.72*
MLR[a]								
Milk wt.[b]	-.119±.080	0.73	-.090±.072	0.54		-.776±.309*	2.29	
% Fat	0.123±.119	0.35	-.008±.131	<0.01		0.129±.228	0.12	
% Protein	0.192±.264	0.17	0.571±.293*	1.29		-.024±.365	<0.01	
% Lactose	-.143±.304	0.07	-.341±.370	0.29		0.220±.421	0.10	
MLR equation		3.96		5.50	8.38*		2.44	1.47

[a]degrees of freedom: SLR; 294, 282 and 272; MLR; 291, 279 and 269 for overall, within periods and within herds respectively.

[b]regression coefficients $x10^{-4}$.

*significant: regression coefficients by t-tests and differences among levels (within subclass models) by F-tests.

[c]R^2 calculated on the total within subclass sums of squares

shipping large amounts of milk were associated with low within herd-period variance of protein percent, but that increased milk shipments by a herd were not significantly associated with changes in the within herd-period variance of protein percent.

Milk fat percent. Milk fat percent was a significant source of variation for the within period simple linear regression model only. The regression coefficient was 0.171 ± .065. This result indicated that herds shipping milk high in milk fat percent were associated with high within herd-period variance of protein percent.

Protein percent. Percent protein was a significant source of variation for the overall, within period and within herd simple linear regression models. The regression coefficients were 0.572 ± .161, 0.512 ± .152 and 0.697 ± .330 respectively. These results indicated that herds shipping milk high in protein percent were associated with high within herd-period variance of protein percent. The results from the analyses of the within herd regression model indicated that an increase in the level of protein in milk shipped by an individual herd was associated with an increase in the within herd-period variance of protein percent.

Lactose percent. Lactose percent was not a significant source of variation for any of the three simple linear regression models.

Overall multiple linear regression. The partial regression coefficients associated with the independent variables of milk shipment weight and percent protein were significantly different from zero by a t-test. The model accounted for 9.33 percent of the total sums of squares of the dependent variable.

Within period multiple linear regression. Only the independent variable of milk shipment weight was a significant source of variation when the within period multiple linear regression model was fitted. The regression coefficient was $(-.244 \pm .068) \times 10^{-4}$. The model accounted for 8.86 percent of the total sums of squares of the dependent variable. Differences in levels were significant by the F-test.

Within herd multiple linear regression. When the multiple regression was computed on a within herd basis none of the independent variables was a significant source of variation. This model accounted for 4.18 percent of the total sums of squares of the dependent variable. Differences between herd means were not significant by the F-test of differences of levels.

Within strata variance of protein percent. The regression coefficients estimated when the within herd-period variance was calculated without stratification were not significantly different from the coefficients estimated with four strata per period (Table 19B).

Within Herd-Period Variance of Percent Lactose

The within herd-period variances of percent lactose, calculated without stratification and with four strata per period were used as dependent variables, after $\log_e$ transformation, in the regression models.

The estimates of the regression coefficients, t-tests of the coefficients and the proportion of the sums of squares (R^2) accounted for by the regression equations are shown in Table 20A for simple linear regressions and multiple linear regression, both overall, within period and within herd for the $\log_e$ of the within herd-period variance of percent lactose. F-tests of the differences among herds and among periods are also shown in Table 20A. The results for the $\log_e$ of the within herd-period variance of percent lactose calculated with four strata per period are shown in Table 20B.

Simple linear regression. Milk shipment weight was not a significant source of variation for any of the three simple linear regression models.

TABLE 20A

SIMPLE (SLR) AND MULTIPLE LINEAR (MLR) REGRESSION COEFFICIENTS FOR THE REGRESSTION OF THE LOGARITHM OF THE WITHIN HERD-PERIOD VARIANCE OF PERCENT LACTOSE ON KILOGRAMS MILK, PERCENT MILK FAT, PROTEIN AND LACTOSE - NO STRATA

	Overall		Within Period			Within Herd		
	b±S.E.	R^2	b±S.E.	R^2 [c]	F	b±S.E.	R^2 [c]	F
SLR[a]								
Milk wt.[b]	-.100±.074	0.61	-.104±.059	1.11	16.49	0.282±.294	0.34	1.22
% Fat	-.233±.071*	3.53	-.111±.058	1.27	15.38*	-1.481±.205*	16.19	3.32*
% Protein	0.040±.167	0.02	-.259±.138	1.24	16.78*	0.854±.347*	2.18	1.56
% Lactose	-.331±.283	0.46	-.733±.285*	2.30	17.04*	-.056±.378	0.01	1.20
MLR								
Milk wt.[b]	-.170±.076*	1.54	-.145±.063*	1.83		-.297±.279	0.34	
% Fat	-.643±.112*	10.07	-.034±.115	0.03		-1.538±.206*	16.77	
% Protein	1.051±.249*	5.43	-.219±.256	0.25		0.920±.330*	2.33	
% Lactose	0.124±.286	0.06	-.493±.325	0.79		-.191±.380	0.08	
MLR equation		11.12		4.51	13.39*		19.56	2.66*

[a] degrees of freedom: SLR; 294, 282 and 272; MLR; 291, 279 and 269 for overall, within periods and within herds respectively.

[b] regression coefficients x10^{-4}.

* significant: regression coefficients by t-tests and differences among levels (within subclass models) by F-tests.

[c] R^2 calculated on the total within subclass sums of squares

TABLE 20B

SIMPLE AND MULTIPLE LINEAR REGRESSION COEFFICIENTS
PERCENT LACTOSE WITH FOUR STRATA PER PERIOD

	Overall		Within Period			Within Herd		
	b±S.E.	R^2	b±S.E.	R^2 [c]	F	b±S.E.	R^2 [c]	F
SLR[a]								
Milk wt.[b]	-.080±.077	0.37	-.091±.065	0.68	10.21*	-.098±.299	0.04	1.46
% Fat	-.179±.073*	1.99	-.091±.065	0.69	9.67*	-1.072±.218*	8.23	2.46*
% Protein	0.120±.172	0.16	-.167±.154	0.41	10.19	0.898±.352*	2.34	1.82*
% Lactose	-.444±.290	0.79	-.450±.320	0.70	10.08*	- .487±.383	0.59	1.48
MLR[a]								
Milk wt.[b]	-.121±.079	0.73	-.128±.071	1.16		0.008±.288	0.01	
% Fat	-.544±.117*	6.87	-.113±.130	0.27		-1.173±.219*	9.44	
% Protein	1.015±.260*	4.83	-.008±.290	0.01		0.825±.351*	1.82	
% Lactose	-.101±.298	0.04	-.186±.367	0.09		-.644±.404	0.83	
MLR equation		8.18		2.12	7.94*		11.93	2.07*

[a]degrees of freedom: SLR; 294, 282 and 272; MLR; 291, 279 and 269 for overall, within periods and within herds respectively.

[b]regression coefficients x10^{-4}.

*significant: regression coefficients by t-tests and differences among levels (within subclass models) by F-tests.

[c]R^2 calculated on the total within subclass sums of squares

Percent milk fat was a significant source of variation when the regression was computed overall and within herds. The regression coefficients were -.233 ± .071 and -1.481 ± .205 respectively. These results indicated that the within herd-period variance of percent lactose increased, for a herd, when the fat content of the milk dropped (spring, see Figure 4).

Percent protein was a significant source of variation for the within herd regression only. The regression coefficient was 0.854 ± .347. This result indicated that the within herd-period variance of percent lactose increased when protein content of herd milk increased (Figure 4).

Percent lactose was a significant source of variation for the within period model only. The regression coefficient was -.733 ± .285 indicating that herds with low lactose levels were significantly higher in the variance of percent lactose.

The simple regression coefficients estimated when the variances were calculated without stratification (Table 20A) were not significantly different from the regression coefficients estimated with four strata per period.

Multiple linear regression. For the overall regression milk shipment weight, percent milk fat and percent protein were all significant sources of variation.

The partial regression coefficients were; $(-.170 \pm .076) \times 10^{-4}$, $-.643 \pm .112$ and $1.051 \pm .249$ respectively. The model accounted for 11.12 percent of the total sums of squares of the dependent variable.

On a within period basis only milk shipment weight was a significant source of variation; the partial regression coefficient was $(-.145 \pm .063) \times 10^{-4}$. The model accounted for 4.51 percent of the total within period sums of squares. Period levels were significantly different by the F-test.

On a within herd basis two of the independent variables, percent milk fat and percent protein, were significant sources of variation. The partial regression coefficients were $-1.538 \pm .206$ and $0.920 \pm .330$ respectively. The model accounted for 19.56 percent of the total within herd sums of squares. The F-test of differences in herd levels was significant.

The regression coefficients estimated when the variances were calculated without stratification were not significantly different from the regression coefficients estimated with four strata per period. The F-test of levels of both periods and herds were significant in both cases.

Conclusion of Regression Analyses

Although the regression analyses showed that in many cases the variances of milk constituent percentages were significantly associated with the independent variables used, the proportion of the total sums of squares accounted for by the regression equations was relatively low and therefore the regression equations have little value for predicting the within herd-period variance of an individual herd-period subclass. The regression analyses also showed that differences among herds and among periods in within herd-period variances of milk constituent percentages were significant.

Herd and Period Variation

The criterion of precision used in the current study was that random sample estimates of herd-period milk constituent percentages should be at least as precise as composite estimates (i.e. that the level of precision of current sampling methods was acceptable to the industry). The variance of estimates that will meet this criterion were presented in Table 14. By rearrangement of equation 24 to yield:

$$\sigma_w^2 = n(N\sigma_{\bar{x}}^2 - \sigma_a^2)/(N - n) \qquad (37)$$

The maximum value of the within herd period variance of milk constituent percentages that will satisfy this criterion can be calculated for a given sample size. The values presented in Table 14 were substituted in equation 37 for $\sigma_{\bar{x}}^2$, values for σ_a^2 (defined in equation 2) were taken from Table 3, to calculate maximum values of within herd-period variance for two, three, four and five random samples per fifteen shipment period for all three milk constituents (Table 21). The values in Table 21 were used to calculate the proportion of herds, periods or individual herd-period subclasses that would meet this criterion for various sampling schemes.

TABLE 21

MAXIMUM VALUE OF σ_w^2 FOR THE PRECISION OF A RANDOM SAMPLE TO MEET THE SPECIFIED CRITERION

Milk Constituent	Variance ($x10^{-2}$)			
	Sample Size			
	Two	Three	Four	Five
% Milk fat	0.737	1.198	1.742	2.395
% Protein	0.280	0.373	0.542	0.745
% Lactose	0.473	0.676	0.983	1.351

Season Variation

The regression analyses showed that differences among periods in the within herd-period variance of milk constituent percent were significant for all milk constituents when the independent variables were held constant. As the data in the current study were collected over only one year no comparisons of season effects across years are possible. If seasons are different estimates of herd-period means would be more precise in some seasons than in others under the same random sampling scheme; therefore, it could be worthwhile to take more samples in some seasons than in others. Alternatively the sampling frequency should be great enough that the criterion of precision is satisfied for the most variable seasons, this would mean that the sampling frequency will be greater in some seasons than necessary. This course of action would be wasteful of resources and would increase costs associated with sampling and testing bulk milk.

Within herd-period variance of milk fat percent. The period variance of milk fat percent was calculated for each period by dividing the pooled within herd sums of squares by the pooled degrees of freedom with no strata and with two, three and four strata per period (Table 22). The results showed that the within herd-period variance of

TABLE 22

PERIOD AVERAGE WITHIN HERD-PERIOD VARIANCE (σ_w^2) OF PERCENT MILK FAT WITHOUT STRATIFICATION AND WITH TWO, THREE AND FOUR STRATA

Period Number	Variance ($x10^{-2}$)							
	No Strata		Two Strata		Three Strata		Four Strata	
	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF
1	1.570±.126	309	1.449±.121	286	0.974±.085	263	1.124±.102	240
2	2.584±.205	315	1.474±.122	292	1.153±.099	269	1.173±.105	246
3	2.291±.182	314	1.518±.125	291	1.374±.118	268	1.257±.113	245
4	0.905±.072	317	0.713±.059	294	0.527±.045	271	0.601±.054	248
5	1.364±.112	293	0.920±.079	266	0.850±.076	247	0.992±.093	224
6	1.563±.126	307	1.380±.115	284	1.276±.111	261	1.052±.096	238
7	1.277±.105	295	1.005±.087	268	0.922±.082	249	0.875±.082	226
8	1.634±.132	304	1.389±.117	281	1.344±.118	258	1.362±.125	235
9	1.159±.093	310	0.734±.061	287	0.608±.053	264	0.619±.056	241
10	1.279±.102	315	1.086±.090	288	1.115±.096	269	1.242±.112	246
11	1.055±.092	261	0.993±.098	204	0.842±.079	227	0.760±.074	207
12	0.614±.049	317	0.505±.041	294	0.509±.044	271	0.441±.039	248
13	0.677±.055	303	0.598±.050	288	0.545±.047	267	0.491±.044	244

percent milk fat was the highest in the spring and early summer (Figure 5). These results agree with those reported by O'Keeffe [16]; however, Herrmann and Anderson [11] and Boswell et al. [3] found that the variance was the highest in the period October to December, although the work of Boswell et al. [3] showed a secondary peak in May. In the current study these values (without stratification), ranged from 0.0258 ± .00205 in the second period (second half of April and the first half of May) to 0.00614 ± .00049 in the twelveth period (end of February and beginning of March). Stratification into four strata resulted in a reduction of the within herd-period variance of milk fat percentage in all periods; however, the reduction was, in general, greater in those periods of high variance than in those periods of low variance (Figure 5). The period variances estimated with four strata per period were all lower than the maximum values shown in Table 21 for four samples per period. Therefore with four samples per period (one from each of four strata) the criterion of precision would be met in all periods. Three samples (possibly two in some months) would be adequate in the winter if the seasonal trends reported in the current study are consistent across years.

The differences among seasons may be due to changes in biological variance or testing variance (Part 1).

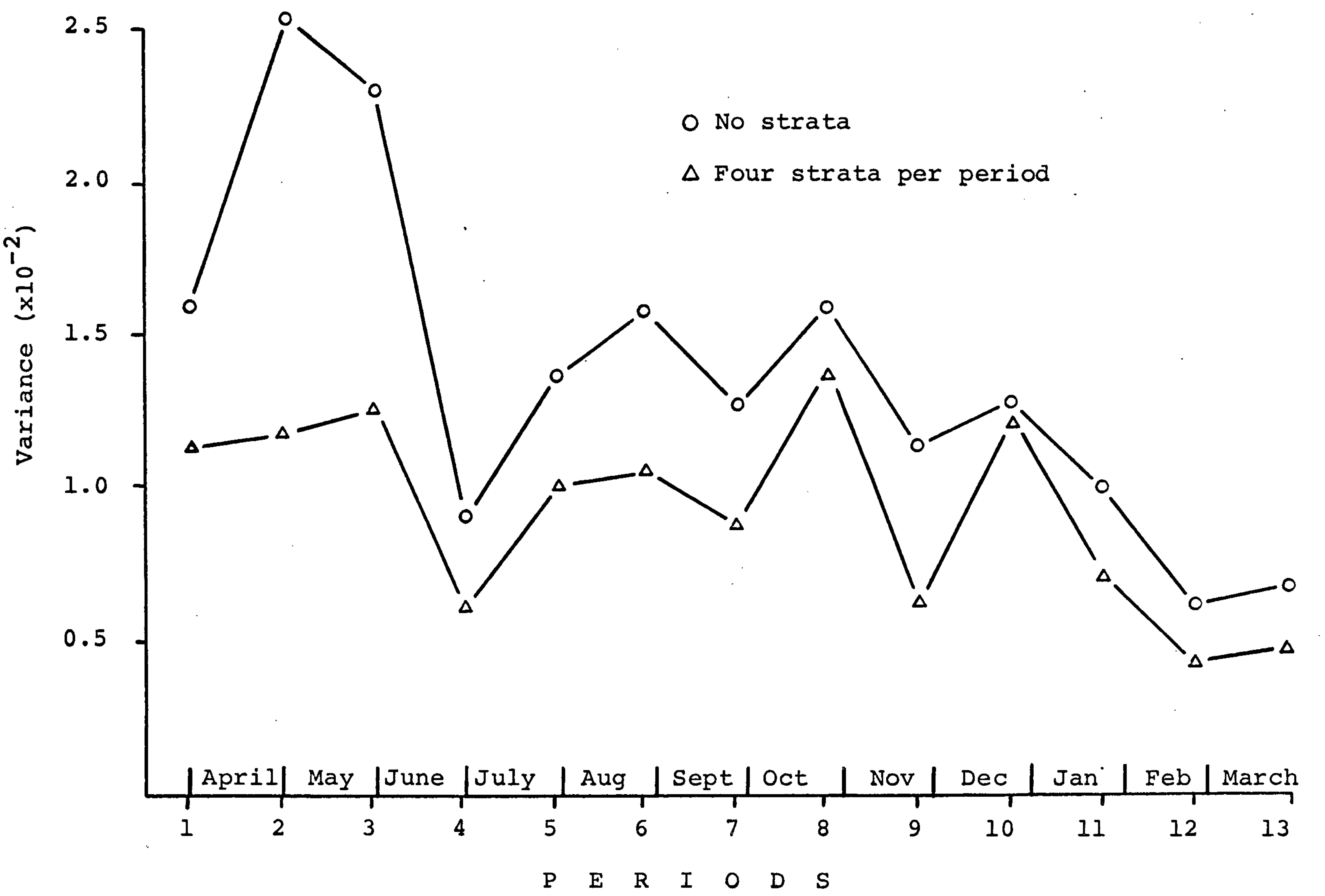

Figure 5 Within herd-period variance of milk fat percent for thirteen periods

However, as within herd-period variance was generally large in those seasons (spring and autumn) associated with changes in herd feeding and handling, the seasonal differences in within herd-period variance can probably be attributed mainly to differences in biological variance.

Within herd-period variance of protein percent. Period means within herd-period variance of percent protein were also calculated with no strata and with two, three and four strata (Table 23). The results (graphed Figure 6) showed two peaks; one in the spring (period two) and one in the autumn (period eight). Stratification resulted in a reduction in the estimates of the within herd-period variance of percent protein in all periods. With four samples per period the variance was higher than the maximum allowable for four of the periods (period two, five, eight and ten). However, as estimates of within herd-period variance of percent protein were lower than that of percent milk fat the standard error of the estimate of percent protein would be lower than the standard error of the estimate of percent milk fat.

Period means within herd-period variance of percent lactose were calculated (Table 24). The results (graphed in Figure 7) showed that with four samples per period the criterion of precision was met in all periods.

TABLE 23

PERIOD AVERAGE WITHIN HERD-PERIOD VARIANCE (σ_w^2) OF PERCENT PROTEIN WITHOUT STRATIFICATION AND WITH TWO, THREE AND FOUR STRATA

	Variance ($\times 10^{-2}$)							
	No Strata		Two Strata		Three Strata		Four Strata	
Period Number	$\sigma_w^2 \pm$S.E.	DF	$\sigma_w^2 \pm$S.E.	DF	$\sigma_w^2 \pm$S.E.	DF	$\sigma_w^2 \pm$S.E.	DF
1	0.536±.043	309	0.509±.042	286	0.504±.044	263	0.433±.039	240
2	1.238±.098	315	1.120±.092	292	1.037±.089	269	0.656±.059	246
3	0.650±.052	314	0.585±.048	291	0.543±.047	268	0.563±.051	245
4	0.714±.057	317	0.453±.037	294	0.388±.033	271	0.310±.028	248
5	0.753±.062	293	0.757±.065	266	0.741±.066	247	0.638±.060	224
6	0.941±.076	307	0.641±.054	284	0.727±.063	261	0.529±.048	238
7	0.723±.059	295	0.596±.051	268	0.482±.043	249	0.373±.035	226
8	1.490±.121	304	1.035±.087	281	0.887±.078	258	1.007±.093	235
9	0.777±.062	310	0.598±.050	287	0.603±.052	264	0.535±.049	241
10	0.887±.070	315	0.790±.066	288	0.740±.064	269	0.808±.073	246
11	0.494±.043	261	0.369±.036	204	0.409±.038	227	0.421±.041	207
12	0.611±.048	317	0.551±.045	294	0.563±.048	271	0.486±.043	248
13	0.565±.046	303	0.450±.037	288	0.325±.028	267	0.263±.024	244

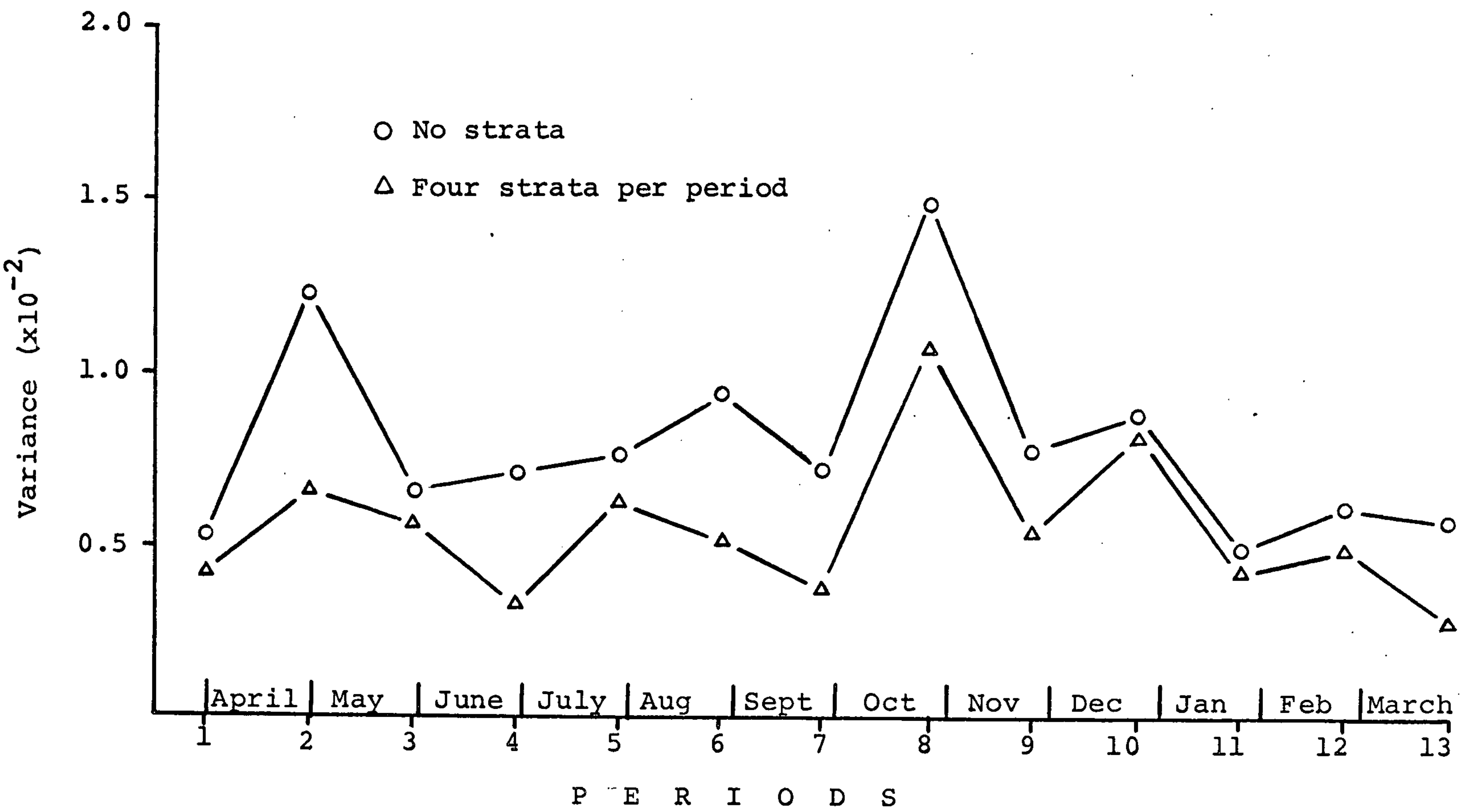

Figure 6 Within herd-period variance of lactose percent for thirteen periods

TABLE 24

PERIOD AVERAGE WITHIN HERD-PERIOD VARIANCE (σ_w^2) OF PERCENT LACTOSE WITHOUT STRATIFICATION AND WITH TWO, THREE AND FOUR STRATA

	Variance ($\times 10^{-2}$)							
	No Strata		Two Strata		Three Strata		Four Strata	
Period Number	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF
1	0.392±.031	309	0.333±.028	286	0.338±.029	263	0.260±.024	240
2	0.822±.065	315	.568±.047	292	.499±.043	269	.414±.037	246
3	.511±.041	314	.519±.043	291	.537±.046	268	.556±.050	245
4	.972±.077	317	.737±.061	294	.638±.055	271	.527±.047	248
5	.525±.043	293	.442±.038	266	.426±.038	247	.452±.043	224
6	.675±.054	307	.535±.045	284	.605±.053	261	.602±.055	238
7	.888±.073	295	.833±.072	268	.821±.073	249	.730±.068	226
8	.541±.044	304	.437±.037	281	.345±.030	258	.407±.037	235
9	.223±.018	310	.215±.018	287	.182±.016	264	.193±.018	241
10	.404±.032	315	.374±.031	288	.271±0.23	269	.366±.033	246
11	.364±.032	261	.278±.027	204	.365±.034	227	.345±.034	207
12	.544±.043	317	.527±.043	294	.513±.044	271	.458±.041	248
13	.253±.020	303	.265±.022	288	.255±.022	267	.248±.022	244

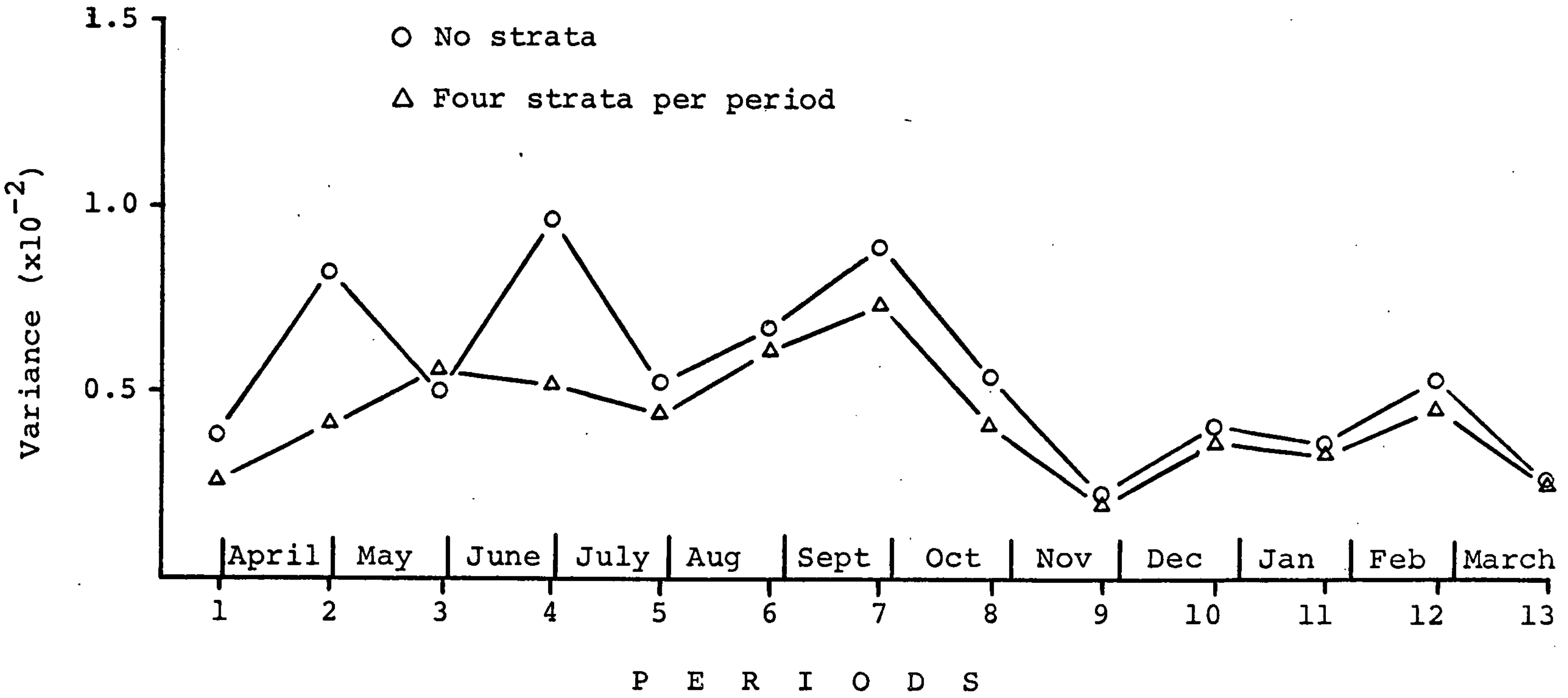

Figure 7 Within herd-period variance of lactose percent for thirteen periods

Herd Variation

The within period regression analyses (Tables 18A to 20B) showed that large herds were lower in within herd-period variance of milk constituent percent than smaller herds and also, in general, that higher variance was associated with high herd levels of milk fat and protein. Herd means of within herd period variance of milk constituent percent were calculated for the twenty-three herds used in the regression analyses with no strata and with two, three and four strata for the within herd-period variance of milk constituent percent.

The within herd-period variance of percent milk fat (Table 25) herd means ranged from 0.02955 ± .00313 for the most variable herd to 0.00590 ± .00063 for the least variable without stratification. With four strata per period the range was from 0.01695 ± .00205 to 0.00380 ± .00046. Therefore with four samples and four strata per period the criterion (Table 21) was met for all herds.

Herd means of within herd-period variance of percent protein (Table 26) ranged from 0.01227 ± .00132 to 0.00458 ± .00049 without stratification and from 0.00968 ± .00117 to 0.00322 ± .00039 with four strata per period. For nearly half the herds the criterion (Table 21) of precision will not be met with four samples (one from each of four strata).

TABLE 25

HERD AVERAGE WITHIN HERD-PERIOD VARIANCE (σ_w^2) OF PERCENT MILK FAT WITHOUT STRATIFICATION AND WITH TWO, THREE AND FOUR STRATA

Variance ($x10^{-2}$)							
No Strata		Two Strata		Three Strata		Four Strata	
σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF
2.955±.313	176	1.402±.154	163	1.215±.139	150	1.453±.174	137
2.081±.221	176	1.858±.207	159	1.714±.197	150	1.540±.185	137
2.020±.215	174	1.624±.182	157	1.728±.200	148	1.695±.205	135
1.889±.202	173	1.741±.193	160	1.484±.172	147	1.579±.192	134
1.675±.184	163	1.137±.130	151	1.053±.124	142	1.182±.146	130
1.667±.185	161	0.904±.102	156	0.774±.091	143	0.801±.098	131
1.580±.168	175	1.370±.151	162	1.039±.120	149	1.173±.141	136
1.526±.164	172	0.962±.109	155	0.818±.095	146	0.632±.077	133
1.443±.156	170	0.854±.097	153	0.669±.078	144	0.602±.074	131
1.423±.156	164	1.103±.123	159	0.964±.111	148	0.996±.120	135
1.418±.152	173	1.312±.147	156	0.847±.098	147	0.888±.108	134
1.313±.139	177	1.024±.112	164	0.901±.103	151	0.909±.109	138
1.244±.133	174	1.128±.127	157	0.867±.100	148	0.988±.119	135

TABLE 25 (continued)

Variance ($x10^{-2}$)							
No Strata		Two Strata		Three Strata		Four Strata	
α_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF
1.201±.130	170	1.138±.129	154	0.997±.117	143	0.907±.111	131
1.165±.125	172	0.983±.111	155	1.003±.117	146	0.939±.114	133
1.152±.123	174	0.792±.088	161	0.750±.087	148	0.688±.083	135
1.117±.119	175	0.912±.102	158	0.759±.087	149	0.757±.091	136
1.005±.109	169	0.977±.111	152	0.837±.098	143	0.726±.089	130
0.979±.103	179	0.811±.088	166	0.725±.082	153	0.732±.087	140
0.937±.100	173	0.599±.067	156	0.549±.064	147	0.508±.062	134
0.889±.095	172	0.705±.080	155	0.649±.075	146	0.536±.065	133
0.664±.071	175	0.581±.065	158	0.552±.063	149	0.562±.068	136
0.590±.063	173	0.420±.047	156	0.371±.043	147	0.380±.046	134

TABLE 26

HERD AVERAGE WITHIN HERD-PERIOD VARIANCE (σ_w^2) OF PERCENT PROTEIN WITHOUT STRATIFICATION AND WITH TWO, THREE AND FOUR STRATA

Variance ($x10^{-2}$)							
No Strata		Two Strata		Three Strata		Four Strata	
σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF
1.227±.132	172	1.067±.120	155	1.008±.117	146	0.853±.104	137
1.199±.127	176	0.856±.095	159	0.846±.097	150	0.672±.081	137
1.175±.128	170	0.878±.099	154	0.964±.113	143	0.793±.097	135
1.075±.118	164	0.826±.092	159	0.845±.097	148	0.725±.088	134
1.060±.113	173	1.064±.118	160	1.083±.125	147	0.969±.117	130
0.882±.094	174	0.608±.067	161	0.491±.057	148	0.511±.062	131
0.876±.096	163	0.674±.077	151	0.727±.086	142	0.561±.069	136
0.872±.093	174	0.778±.087	157	0.695±.080	148	0.668±.081	133
0.864±.091	179	0.764±.083	166	0.720±.082	153	0.666±.079	131
0.791±.086	169	0.520±.059	152	0.436±.051	143	0.343±.042	135
0.791±.084	175	0.658±.073	162	0.529±.061	149	0.476±.057	134
0.776±.082	175	0.557±.062	158	0.465±.062	149	0.426±.051	138
0.752±.083	161	0.581±.065	156	0.542±.064	143	0.497±.061	135

TABLE 26 (continued)

Variance ($x10^{-2}$)							
No Strata		Two Strata		Three Strata		Four Strata	
$\sigma_W^2 \pm$S.E.	DF	$\sigma_W^2 \pm$S.E.	DF	$\sigma_W^2 \pm$S.E.	DF	$\sigma_W^2 \pm$S.E.	DF
0.748±.080	172	0.604±.068	155	0.538±.063	146	0.436±.053	131
0.732±.078	176	0.664±.073	163	0.579±.066	150	0.561±.067	133
0.712±.076	174	0.637±.071	157	0.605±.070	148	0.572±.069	135
0.702±.075	173	0.427±.048	156	0.455±.053	147	0.378±.045	136
0.693±.074	175	0.597±.067	158	0.563±.065	149	0.534±.064	130
0.599±.064	172	0.499±.056	155	0.405±.047	146	0.329±.040	140
0.537±.057	173	0.503±.057	156	0.477±.055	147	0.376±.046	134
0.478±.051	173	0.427±.048	156	0.339±.039	147	0.322±.039	133
0.464±.049	177	0.426±.047	164	0.361±.041	151	0.365±.044	136
0.458±.049	170	0.459±.052	153	0.429±.050	144	0.395±.048	134

The within herd-period variance of percent lactose (Table 27) herd means ranged from 0.00836 ± .00089 to 0.00352 ± .00038 without stratification and from 0.00714 ± .00086 to 0.00276 ± .00034 with four strata per period. The herd means are all below the maximum value allowed if the criterion of precision (Table 21) is to be met and four samples are taken each period.

Laboratory determinations were done for all herds at approximately the same time; therefore differences among herds can mainly be attributed to differences in biological variance.

Distribution of Within Herd-Period Variances

Table 28 shows the frequency distribution of the within herd-period variances of milk fat percent calculated with no strata and with two, three and four strata per period. A histogram of the distribution is shown in Figure 8 for no strata and for four strata. With four samples taken at random in a period (no strata) 77.57 percent (Table 31) of the individual herd-periods were predicted to meet the specified criterion of precision (Table 14). With three stratified random samples (one sample from each of three strata) 77.14 percent of the herd-periods will also meet the same standard; therefore, stratification will result in the saving of one sample

TABLE 27

HERD AVERAGE WITHIN HERD-PERIOD VARIANCE OF PERCENT LACTOSE WITHOUT STRATIFICATION AND WITH TWO, THREE AND FOUR STRATA

Variance ($x10^{-2}$)							
No Strata		Two Strata		Three Strata		Four Strata	
$\sigma_w^2 \pm$S.E.	DF	$\sigma_w^2 \pm$S.E.	DF	$\sigma_w^2 \pm$S.E.	DF	$\sigma_w^2 \pm$S.E.	DF
0.836±.089	175	0.786±.087	162	0.744±.086	149	0.714±.086	137
.806±.087	170	.763±.086	154	.615±.072	143	.618±.076	137
.726±.077	175	.564±.063	158	.532±.061	149	.522±.063	135
.681±.072	176	.622±.069	159	.630±.072	150	.574±.069	134
.616±.066	172	.482±.054	155	.438±.051	146	.416±.051	130
.609±.065	175	.590±.065	163	.574±.066	150	.582±.070	131
.599±.064	175	.525±.059	158	.470±.054	149	.459±.055	136
.581±.061	177	.521±.057	164	.443±.051	151	.447±.053	133
.580±.062	172	.446±.050	155	.460±.054	146	.420±.051	131
.540±.060	161	.460±.052	156	.440±.052	143	.417±.051	135
.537±.057	173	.449±.050	160	.449±.052	147	.437±.053	134
.510±.054	174	.413±.046	161	.398±.046	148	.372±.045	138
.507±.056	164	.450±.050	159	.423±.049	148	.377±.046	135

TABLE 27 (continued)

Variance ($x10^{-2}$)							
No Strata		Two Strata		Three Strata		Four Strata	
σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF	σ_w^2±S.E.	DF
.503±.054	173	.412±.046	156	.441±.051	147	.389±.047	131
.501±.054	173	.331±.037	156	.316±.037	147	.276±.034	133
.486±.054	163	.433±.049	151	.362±.043	142	.432±.053	135
.473±.050	174	.379±.042	157	.365±.042	148	.341±.041	136
.467±.051	169	.384±.044	152	.366±.043	143	.345±.042	130
.446±.047	179	.411±.045	166	.384±.044	153	.396±.047	140
.434±.046	174	.371±.042	157	.401±.046	148	.368±.044	134
.420±.045	172	.349±.039	155	.346±.040	146	.342±.042	133
.420±.045	173	.330±.037	156	.333±.039	147	.280±.034	136
.352±.038	170	.312±.035	153	.300±.035	144	.282±.035	134

TABLE 28

FREQUENCY DISTRIBUTION OF THE VARIANCE OF PERCENT MILK FAT CALCULATED WITHOUT STRATIFICATION AND WITH TWO, THREE AND FOUR STRATA PER PERIOD

Relative and Cumulative Frequencies

Number of Strata

Class Limits	None		Two		Three		Four	
	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]
0.0 - 0.0049	17.95		24.52		32.06		33.33	
.005 - .0099	31.73	49.68	40.13	64.65	38.41	70.47	35.56	68.89
.010 - .0149	21.47	71.15	15.92	80.57	14.60	85.07	15.87	84.76
.0150 - .0199	10.90	82.05	8.60	89.17	6.03	91.10	6.98	91.74
.0200 - .0249	5.13	87.18	4.46	93.63	3.81	94.91	3.17	94.91
.0250 - .0299	5.13	92.31	2.87	96.50	2.22	97.13	2.54	97.45
.0300 - .0349	2.56	94.87	1.59	98.09	1.90	99.03	0.32	97.77
.0350 - .0399	1.28	96.15	0.0	98.09	0.0	99.03	1.59	99.36
.0400 - .0449	0.96	97.11	0.0	98.09	0.32	99.35	0.0	99.36
.0450 - .0499	0.96	98.07	1.27	99.36	0.0	99.35	0.32	99.68
.0500 - .0549	0.32	98.39	0.0	99.36	0.32	99.67	0.0	99.68
.0550 - .0599	0.32	98.71	0.32	99.68	0.0	99.67	0.0	99.68
.0600 - .0649	0.32	99.03	0.0	99.68	0.0	99.67	0.32	100.00
.0650 -	0.96	99.99	0.32	100.00	0.32	99.99		

TABLE 28 (continued)

Relative and Cumulative Frequencies

Number of Strata

	None	Two	Three	Four
Mean	0.01371	0.01052	0.00928	0.00926
Stan. Dev.	.01373	.00897	.00785	.00796
Largest Value	.1592	.0685	.06503	.06358
Smallest Value	.00146	.00065	.00133	.00089
Number	312	314	315	315

[a]Cumulative frequencies.

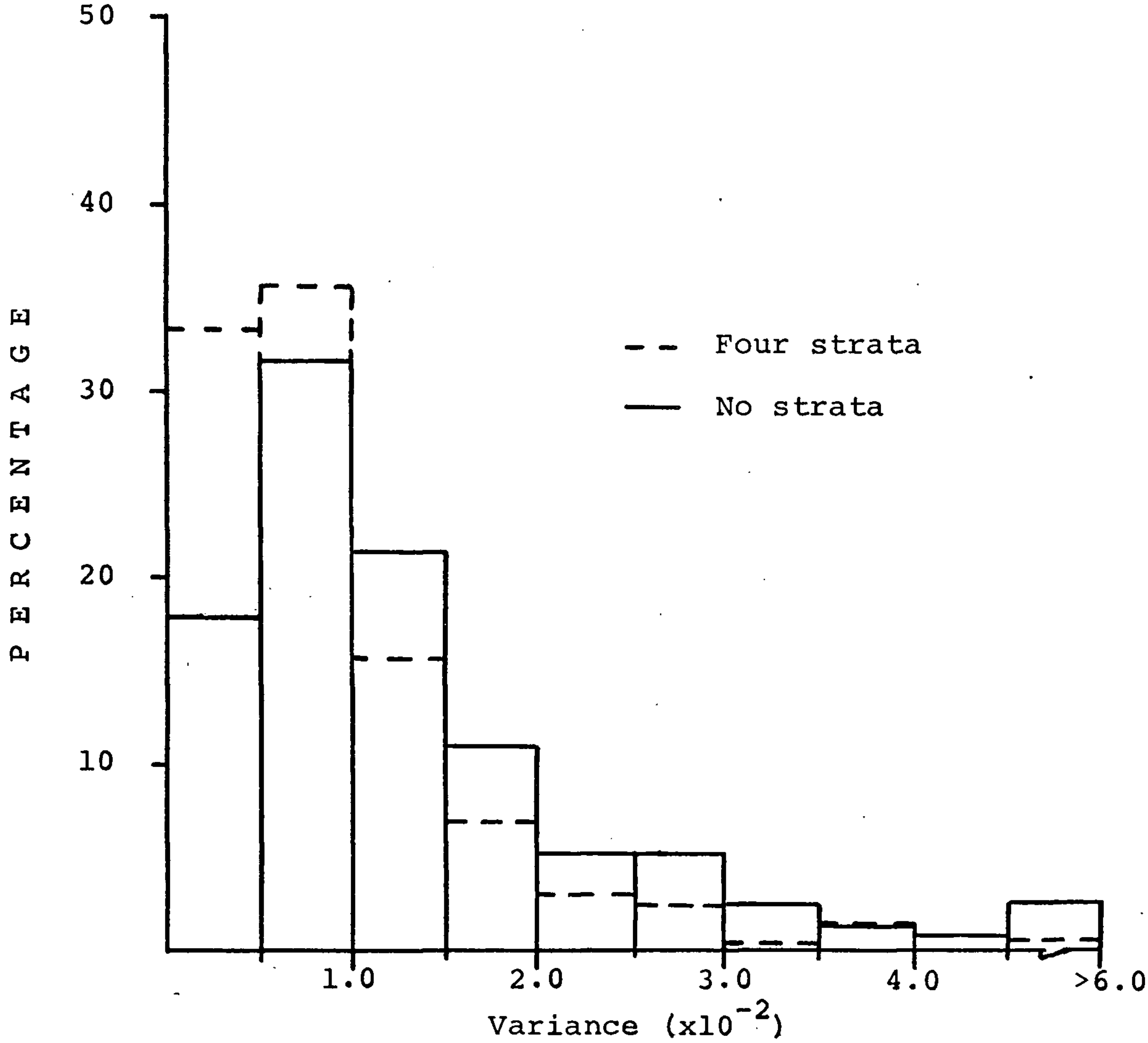

Figure 8 Distribution of the within herd-period variance of milk fat percent (no strata and four strata)

per period in order to meet the same criterion for the same proportion of herd-periods. With four stratified random samples per period 89.53 percent of the subclasses will meet the criterion.

Table 29 shows the frequency distribution of the individual herd-period variances of percent protein without strata and with two, three and four strata per period. A histogram is presented in Figure 9. With four simple random samples per period 41.03 percent of the herd-periods were below the limits specified in Table 21. The percentages for stratified random sampling were; 33.97 and 68.89 for three and four strata respectively (Table 31).

Table 30 shows the frequency distribution of the within herd-period variances of percent lactose with no strata and with two, three and four strata per period. A histogram is presented in Figure 10. With four simple random samples per period 89.42 percent of the herd-period were below the limits specified in Table 21. With stratified random sampling the percentages were 83.81 and 95.87 for three and four strata per period respectively.

All Possible Samples for Seven Sampling Schemes - Experiment I

All possible samples for seven random sampling schemes (Material and Methods) were computer generated from the data of Experiment I. The deviation of each

TABLE 29

FREQUENCY DISTRIBUTION OF THE VARIANCE OF PERCENT PROTEIN CALCULATED WITHOUT STRATA AND WITH TWO, THREE AND FOUR STRATA PER PERIOD

	Relative and Cumulative Frequencies							
	Number of Strata							
	None		Two		Three		Four	
Class Limits	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]
0.0 - 0.0049	37.18		48.41		55.24		64.13	
.0050 - .0099	39.10	76.28	39.17	87.58	32.70	87.94	27.30	91.43
.0100 - .0149	13.14	89.42	6.37	93.95	6.35	94.29	3.49	94.92
.0150 - .0199	6.41	95.83	3.50	97.45	3.49	97.78	3.17	98.09
.0200 - .0249	1.92	97.75	1.27	98.72	1.27	99.05	0.63	98.72
.0250 - .0299	1.28	99.03	0.32	99.04	0.32	99.37	0.95	99.67
.0300 - .0349	0.64	99.67	0.96	100.00	0.32	99.69	0.0	99.67
.0350 - .0399	0.32	99.99			0.32	100.01	0.32	99.99
Mean	.00790		0.00644		0.00600		0.00528	
Stan. Dev.	.00581		.00480		.00485		.00452	
Largest Value	.03954		.03261		.03700		.03655	
Smallest Value	.00061		.00053		.00054		.00059	
Number	312		314		315		315	

[a]Cumulative frequencies.

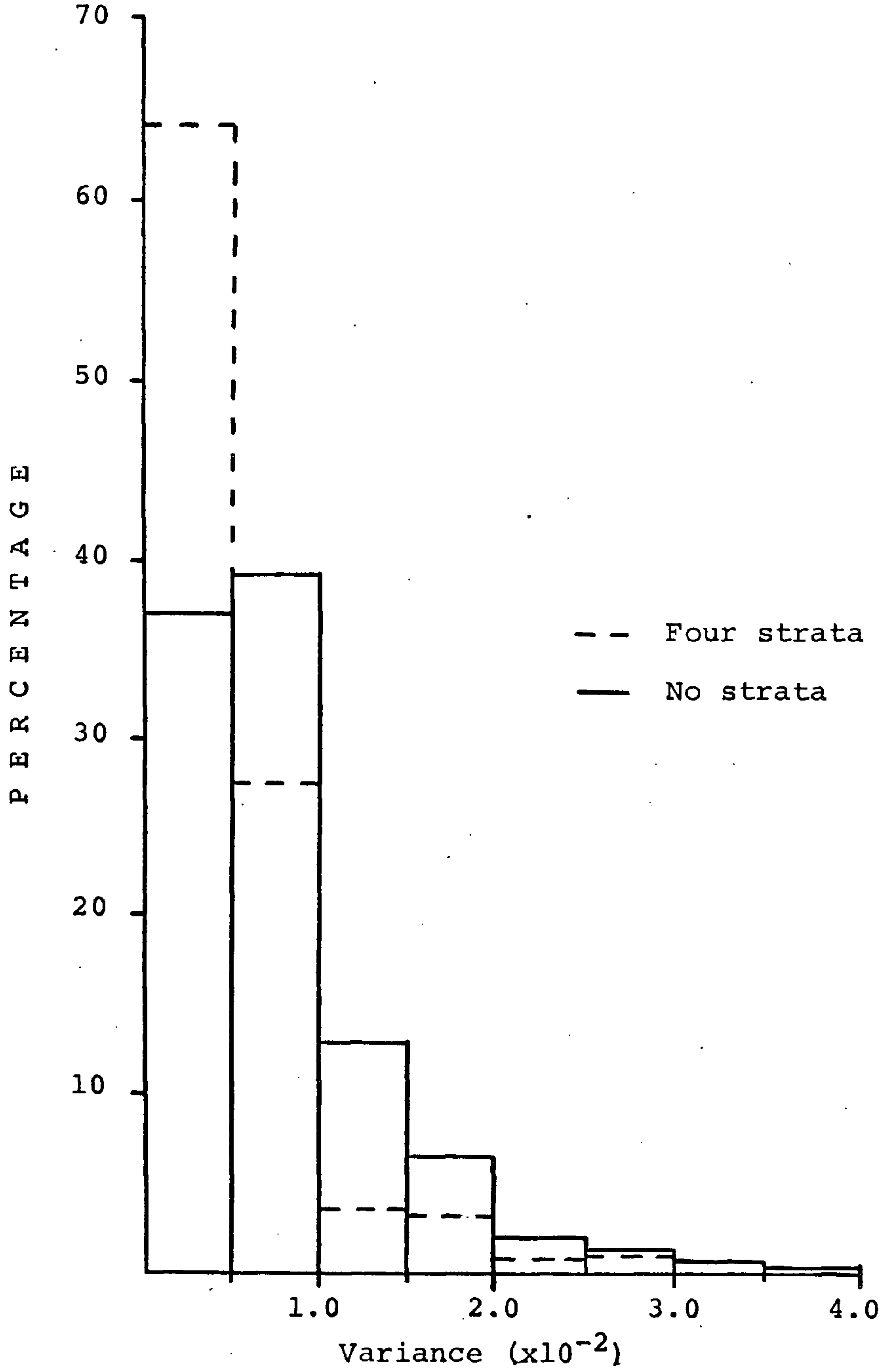

Figure 9 Distribution of the within herd-period variance of protein percent (no strata and four strata)

TABLE 30

FREQUENCY DISTRIBUTION OF THE VARIANCE OF PERCENT LACTOSE CALCULATED WITHOUT STRATA AND WITH TWO, THREE AND FOUR STRATA PER PERIOD

	Relative and Cumulative Frequencies							
	Number of strata							
	None		Two		Three		Four	
Class Limits	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]
0.0 - 0.0049	57.69		68.79		67.94		70.48	
.0050 - .0099	31.73	89.42	25.16	93.95	27.94	95.88	25.40	95.88
.0100 - .0149	7.05	96.47	3.18	97.13	1.59	97.47	3.17	99.05
.0150 - .0199	2.56	99.03	2.23	99.36	1.90	99.37	0.0	99.05
.0200 - .0249	0.32	99.35	0.0	99.36	0.0	99.37	0.32	99.37
.0250 - .0299	0.32	99.67	0.0	99.36	0.32	99.69	0.67	100.00
.0300 - .0349	0.32	99.99	0.64	100.00	0.32	100.01		
Mean	0.00552		0.00470		0.00447		0.00427	
Stan. Dev.	.00412		.00372		.00363		.00331	
Largest value	.03061		.03195		.03325		.02788	
Smallest value	.00054		.00027		.00026		.00028	
Number	312		314		315		315	

[a]Cumulative frequencies.

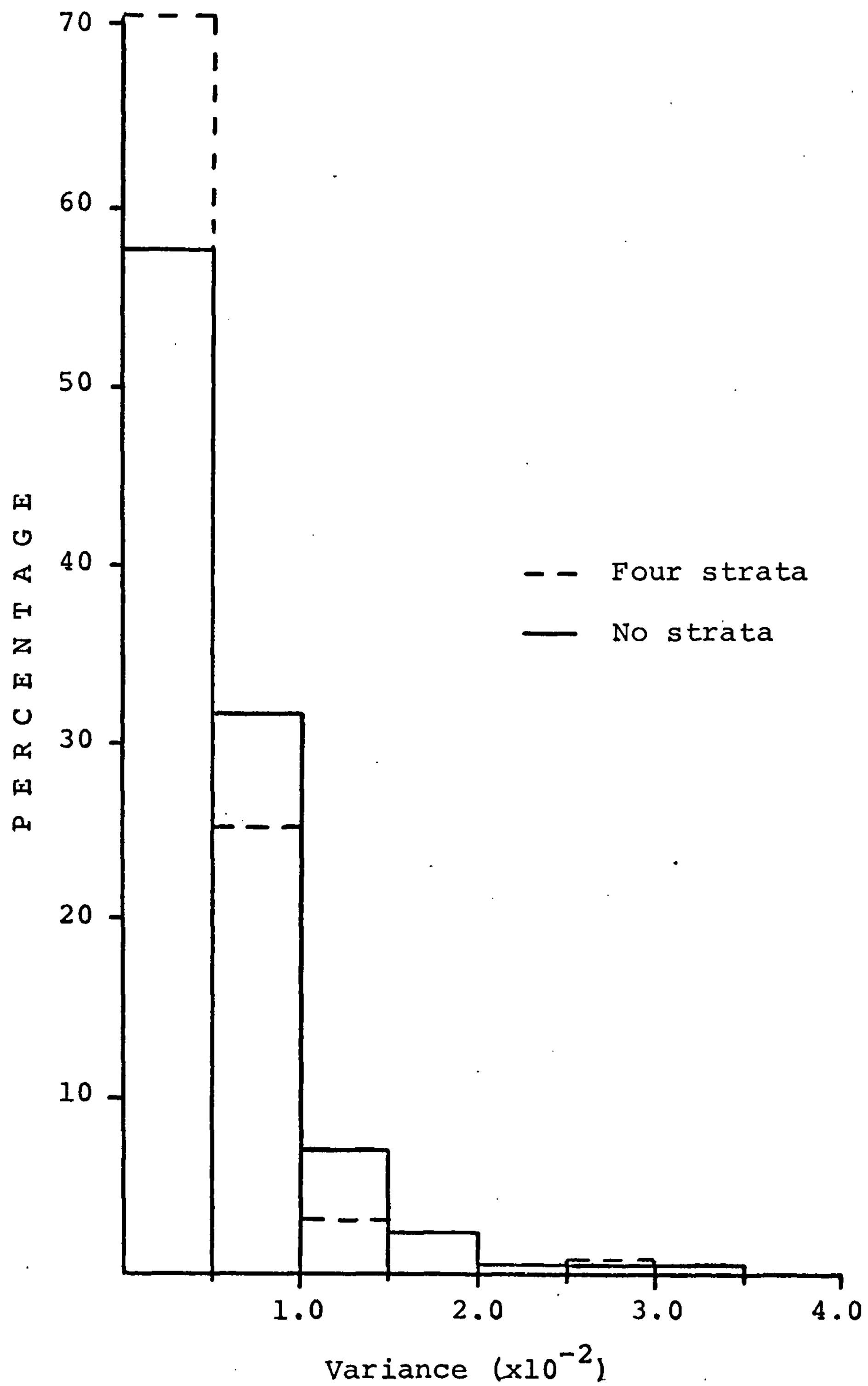

Figure 10 Distribution of the within herd-period variance of lactose percent (no strata and four strata)

TABLE 31

PERCENTAGE OF HERD-PERIOD SUBCLASSES PREDICTED TO MEET THE CRITERION OF PRECISION (TABLE 21)

Number of Samples	Number of Strata		
	None	Three	Four
Percent milk fat			
3	60.26	77.14	
4	77.57		89.53
Percent protein			
3	19.87	33.97	
4	41.03		68.89
Percent lactose			
3	73.08	83.81	
4	89.43		95.87

sample mean (percent milk fat and protein) from the fresh estimate of the herd-period mean were calculated for all possible samples in each herd-period subclass for all herd-periods. Frequency distributions of the absolute value of these deviations were constructed for each of the seven sampling schemes. These frequency distributions indicated the expected results if random sampling had been used for the herds in Experiment I.

The relative and cumulative frequencies of the absolute deviations, the standard error of the mean, the largest absolute deviation and the number of all possible samples for each of the seven sampling schemes are shown in Tables 32A and 32B for percent milk fat and in Tables 33A and 33B for percent protein. Histograms are presented in Figures 11 to 14 of the distribution of the deviations.

The confidence limits for the mean percent milk fat for the seven schemes (Tables 32A and 32B) agree well with those predicted from experiment two (Table 9). The confidence limits for percent protein are larger (Tables 33A and 33B) than predicted (Table 9); however the confidence limits for percent protein are smaller for each of the schemes than for percent milk fat. Therefore, for any sampling scheme, the mean percent protein would be more precisely estimated than the mean percent milk fat. Stratification reduced the frequency of large deviations and the magnitude of the largest deviation. For example: with three samples and three strata the

TABLE 32A

FREQUENCY DISTRIBUTION OF THE ABSOLUTE DEVIATIONS OF ALL POSSIBLE SIMPLE RANDOM SAMPLES, WITH ONE TO FOUR OBSERVATIONS PER SAMPLE, FROM THE PERCENT MILK FAT FRESH MEAN

	Relative and Cumulative Frequencies							
	Number of Observations per Sample							
	One		Two		Three		Four	
Class Limits	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]
0.00 - 0.019	16.43		24.02		30.22		35.78	
0.020 - 0.039	15.45	31.88	21.19	45.21	25.03	55.25	27.48	63.26
0.040 - 0.059	14.08	45.96	17.31	62.52	17.73	72.98	17.26	80.52
0.060 - 0.079	12.21	58.17	12.49	75.01	11.20	84.18	9.47	89.99
0.080 - 0.099	10.00	68.17	8.79	83.80	6.74	90.92	4.91	94.90
0.100 - 0.119	7.47	75.64	5.60	89.40	3.93	94.85	2.49	97.39
0.120 - 0.139	6.01	81.65	3.68	93.08	2.18	97.03	1.21	98.60
0.140 - 0.159	4.72	86.37	2.42	95.50	1.19	98.22	0.64	99.24
0.160 - 0.179	3.44	89.81	1.51	97.01	0.72	98.94	0.32	99.56
0.180 - 0.199	2.86	92.67	1.04	98.05	0.40	99.34	0.18	99.74
0.200 - 0.249	3.95	96.62	1.23	99.28	0.43	99.77	0.18	99.92
0.250 - 0.299	1.80	98.42	0.42	99.70	0.15	99.92	0.05	99.97

TABLE 32A (continued)

Relative and Cumulative Frequencies								
Number of Observations per Sample								
	One		Two		Three		Four	
	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]
.300 -	1.59	100.01	0.30	100.00	0.08	100.00	0.03	100.00
Largest Value	0.742		0.623		0.605		0.538	
Mean Deviation	0.085		0.058		0.038		0.045	
Stan. Dev.[b]	0.113		0.077		0.060		0.050	
99% C.L.[c]	±.292		±.199		±.156		±.129	
No. of Samples	4,511		30,154		124,803		357,659	

[a]Cumulative relative frequencies.

[b]Standard deviation of the distribution of deviations.

[c]99% confidence limits.

TABLE 32B

FREQUENCY DISTRIBUTION OF THE ABSOLUTE DEVIATIONS OF ALL POSSIBLE STRATIFIED RANDOM SAMPLES, WITH ONE OBSERVATION PER STRATA AND TWO TO FOUR STRATA, FROM THE PERCENT MILK FAT FRESH MEAN

	Relative and Cumulative Frequencies					
	Number of Strata					
	Two		Three		Four	
Class Limits	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]
0.00 - 0.019	26.33		34.92		40.61	
.020 - .039	22.78	49.11	27.45	62.37	29.72	70.33
.040 - .059	18.76	67.87	17.87	80.24	16.18	86.51
.060 - .079	11.79	79.66	9.92	90.16	7.57	94.08
.080 - .099	8.16	87.82	5.01	95.17	3.42	97.50
.100 - .119	4.85	92.67	2.49	97.66	1.39	98.89
.120 - .139	2.89	95.56	1.21	98.87	0.65	99.54
.140 - .159	1.86	97.42	0.52	99.39	0.28	99.82
.160 - .179	0.89	98.31	0.30	99.69	0.09	99.91
.180 - .199	0.73	99.04	0.14	99.83	0.05	99.96
.200 - .249	0.65	99.69	0.13	99.96	0.04	100.00
.250 - .299	0.20	99.89	0.04	100.00		
.300 -	0.11	100.00				

TABLE 32B (continued)

	Number of Strata		
	Two	Three	Four
Largest Deviation	0.465	0.295	0.236
Mean Deviation	.051	.038	.032
Stan. Dev.[b]	.0675	.0529	.0416
99% C.L.	.174	.136	.117
Number of Samples	16,081	34,240	50,469

[a]Cumulative relative frequencies.

[b]Standard deviation of the distribution of deviations.

[c]99% confidence limits.

TABLE 33A

FREQUENCY DISTRIBUTION OF THE ABSOLUTE DEVIATIONS OF ALL POSSIBLE SIMPLE RANDOM SAMPLES, WITH ONE TO FOUR OBSERVATIONS PER SAMPLE, FROM THE PERCENT PROTEIN FRESH MEAN

	Relative and Cumulative Relative Frequencies							
	Number of Observations per Sample							
	One		Two		Three		Four	
Class Limits	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]	%	Cum.[a]
0.0 - 0.019	21.41		30.07		36.95		43.23	
.020 - .039	19.53	40.94	25.25	55.32	28.73	65.68	30.44	73.67
.040 - .059	16.47	57.41	18.30	73.62	17.32	83.00	15.59	89.26
.060 - .079	12.28	69.69	11.40	85.02	9.05	92.05	6.47	95.73
.080 - .099	9.47	79.16	6.71	91.73	4.25	96.30	2.60	98.33
.100 - .119	7.27	86.43	3.75	95.48	1.93	98.23	1.04	99.37
.120 - .139	4.61	91.04	1.94	97.42	0.90	99.13	0.40	99.77
.140 - .159	3.08	94.12	1.08	98.50	0.49	99.62	0.15	99.92
.160 - .179	1.97	96.09	0.59	99.09	0.22	99.84	0.05	99.97
.180 - .199	1.42	97.51	0.34	99.43	0.10	99.94	0.02	99.99
.200 - .249	1.42	98.93	0.41	99.84	0.06	100.00	0.01	100.00
.250 - .299	0.53	99.46	0.13	99.97	0.01	100.01		
.300 -	0.53	99.99	0.03	100.00				

TABLE 33A (continued)

	Number of Observations per Sample			
	One	Two	Three	Four
Largest Deviation	0.598	0.447	0.305	0.249
Mean Deviation	.064	.044	.035	.029
Stan. Dev.[b]	.0856	.0585	.0460	.0381
99% C.L.[c]	.221	.151	.119	.099
Number of Samples	4,511	30,154	124,803	357,659

[a]Cumulative relative frequencies.

[b]Standard deviation of the distribution of deviations.

[c]99% confidence limits.

TABLE 33B

FREQUENCY DISTRIBUTION OF THE ABSOLUTE DEVIATIONS OF ALL POSSIBLE STRATIFIED RANDOM SAMPLES, WITH ONE OBSERVATION PER STRATA AND TWO TO FOUR STRATA, FROM THE PERCENT PROTEIN FRESH MEAN

Class Limits		Relative and Cumulative Relative Frequencies — Number of Strata: Two		Three		Four	
0.0	0.019	32.36		41.39		50.71	
.020	.039	26.85	59.21	30.47	71.86	30.97	81.68
.040	.059	18.31	77.52	16.07	87.93	12.56	94.24
.060	.079	11.36	88.88	7.05	94.98	3.80	98.04
.080	.099	5.67	94.55	2.82	97.80	1.18	99.22
.100	.119	2.64	97.19	1.12	98.92	0.50	99.72
.120	.139	1.20	98.39	0.57	99.49	0.19	99.91
.140	.159	0.66	99.05	0.30	99.79	0.07	99.98
.160	.179	0.30	99.35	0.12	99.91	0.01	99.99
.180	.199	0.25	99.60	0.05	99.96	0.01	100.00
.200	.249	0.24	99.84	0.02	99.98		
.250	.299	0.12	99.96	0.01	99.99		
.300		0.03	99.99				

TABLE 33B (continued)

	Number of Strata		
	Two	Three	Four
Largest Deviation	0.377	0.309	0.182
Mean Deviation	.040	.031	.024
Stan. Dev.[b]	.0531	.0404	.0317
99% C.L.[c]	.137	.104	.082
Number of Samples	16,081	34,240	50,469

[a]Cumulative relative frequencies.

[b]Standard deviation of the distribution of deviations.

[c]99% confidence limits.

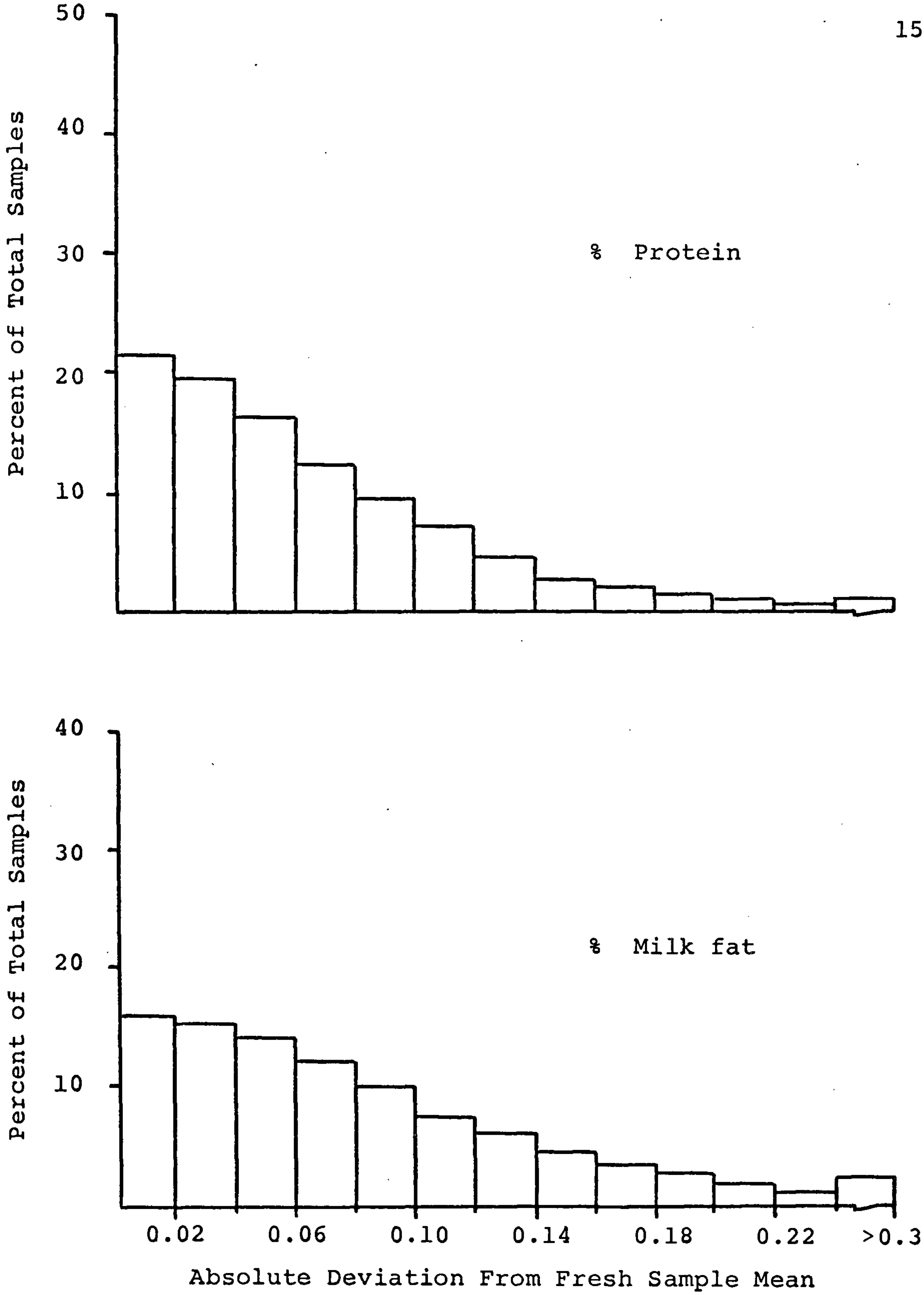

Figure 11 Distribution of absolute deviations of all possible single samples (n=1) from the fresh sample estimate-percent milk fat and protein

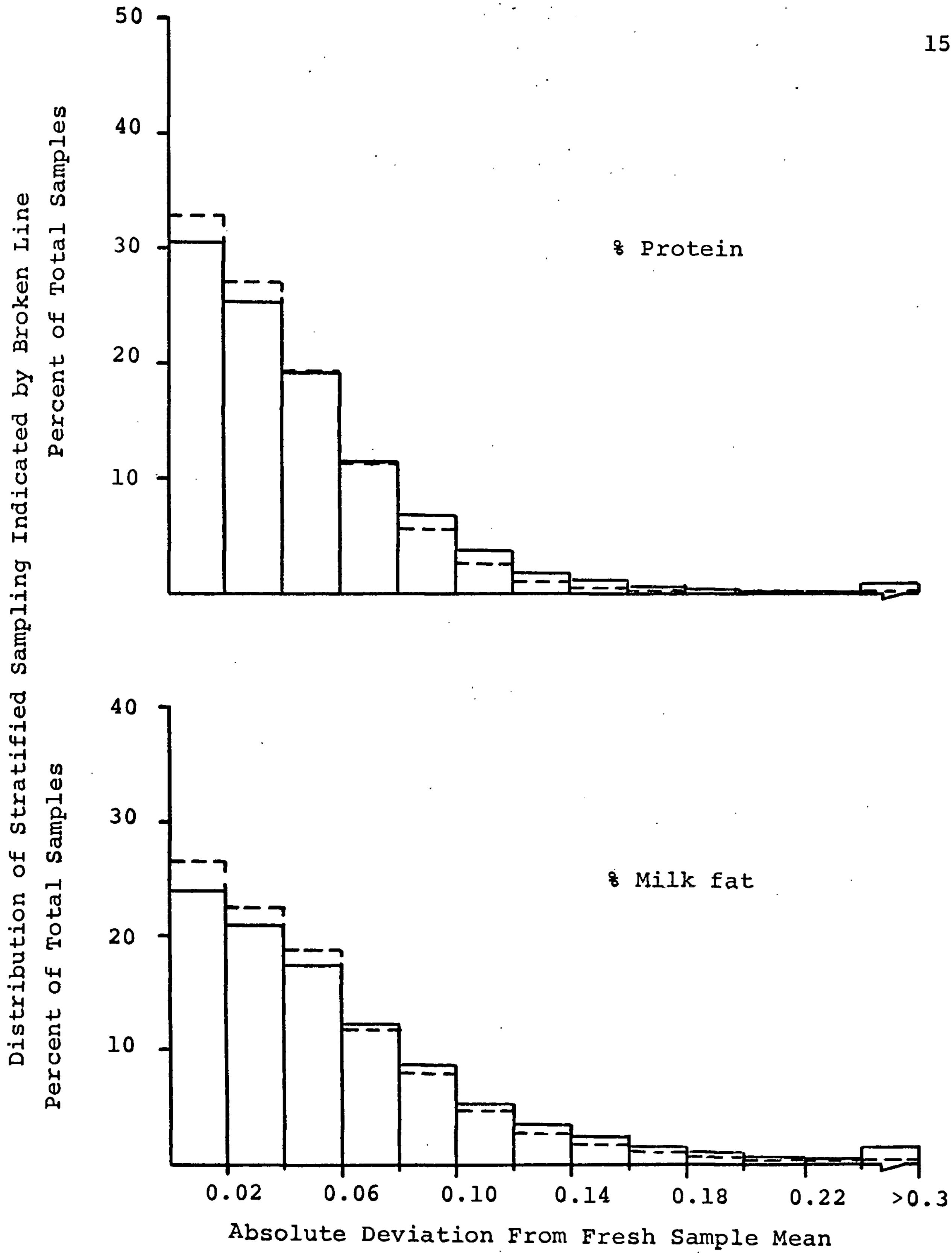

Figure 12 Distribution of absolute deviations of all possible samples of size two (n=2) from the fresh sample estimate-percent milk fat and protein

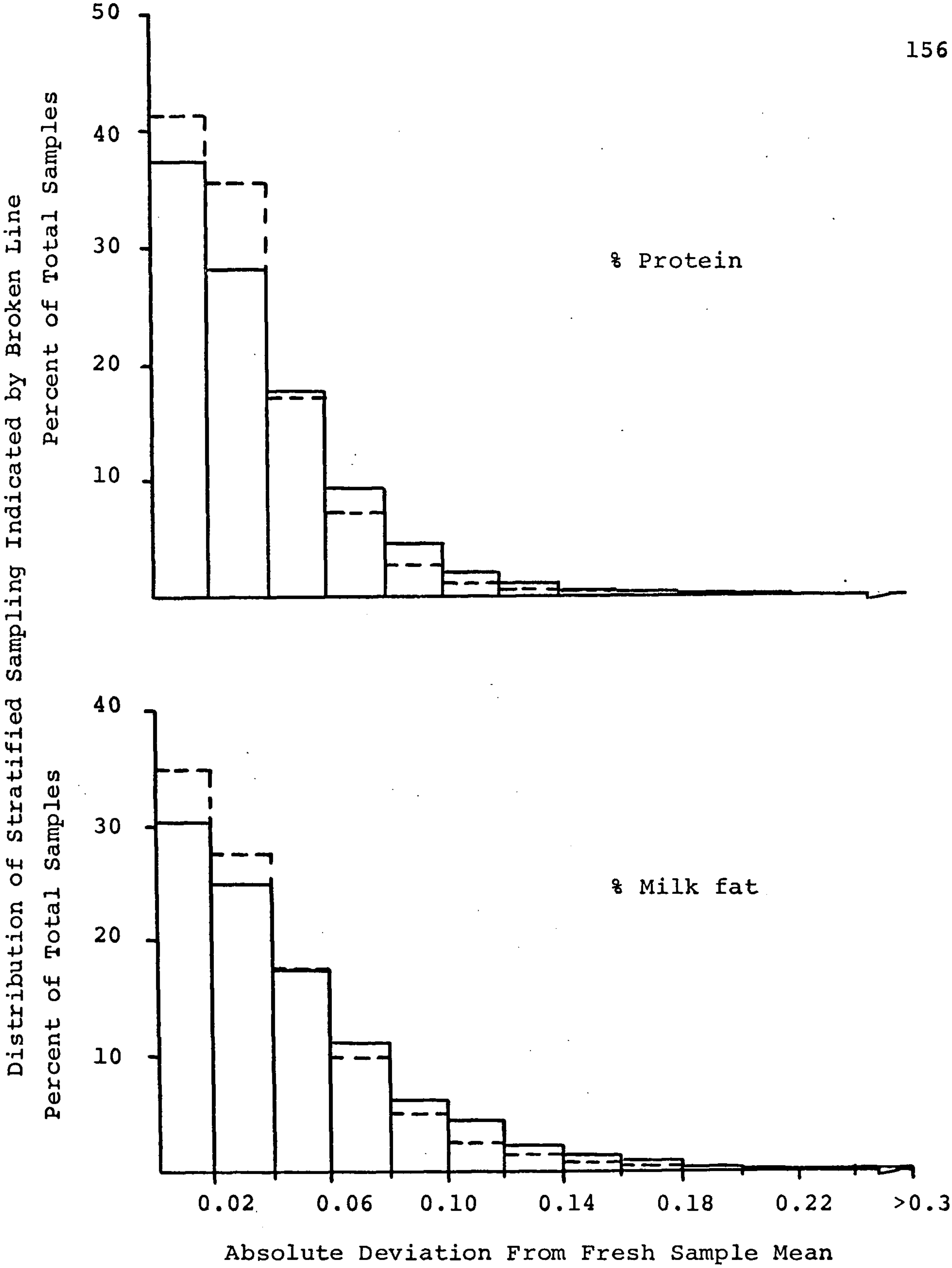

Figure 13 Distribution of absolute deviations of all possible samples of size three (n=3) from the fresh sample estimate-percent milk fat and protein

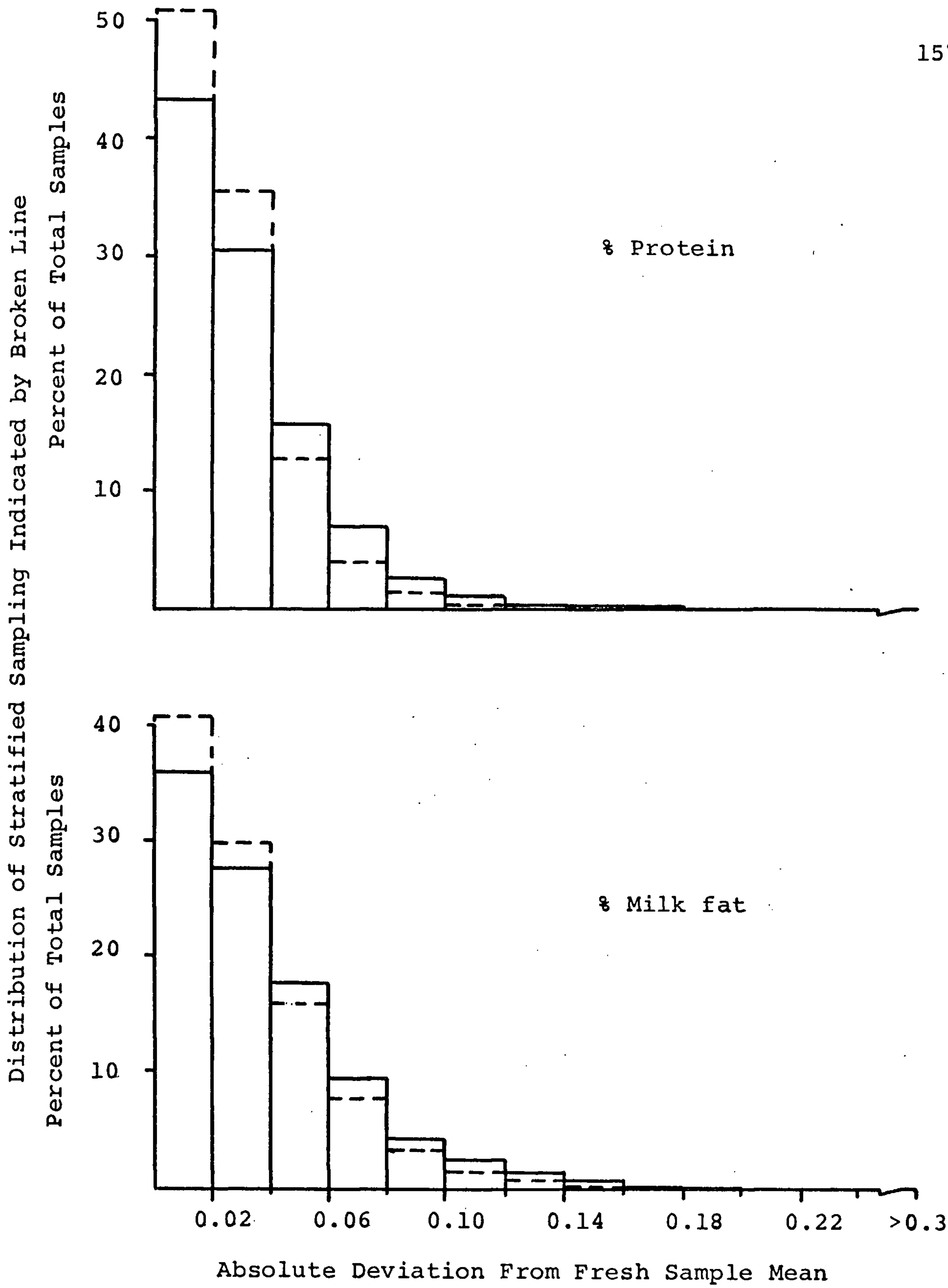

Figure 14 Distribution of absolute deviations of all possible samples of size four (n=4) from the fresh sample estimate-percent milk fat and protein

largest deviation was 0.295 in absolute value but the largest deviation with three simple random samples was 0.605 for percent milk fat; with four samples the largest values were 0.236 and 0.538 percent milk fat for stratified and simple random sampling respectively. In both cases the largest deviation from the fresh mean with stratified sampling was less than one-half as large as the largest deviation with simple random sampling.

Monitoring Random Sampling

A milk sampling scheme should contain provisions for resolving a disputed result (i.e. the producer considers that a particular estimate is too low). As producers receive the results of the analyses after the period to which it applies is over, any additional samples taken in order to settle a disputed result are from milk shipped in the next period and consequently are an unsatisfactory check of the estimate of the previous period mean. Therefore it would be worthwhile to monitor the observed test results as they are accumulated so that the decision to eliminate or replace observations which show large deviations from prior tests could be made before the period is over.

With stratified random sampling (one observation per strata) differences between consecutive milk samples can be attributed to three sources:

1. technical errors such as; sample misidentification, equipment malfunction, etc.
2. errors, in the statistical sense, due to sampling from adjacent strata with the same means and variances.
3. to true but unknown differences between adjacent strata means.

Large deviations between consecutive milk samples due to points 2 and 3 above are expected to occur but are valid unbiased estimates of the true mean and in general the observations should not be replaced or eliminated. Large deviations due to point one above however should be detected and the offending observation should be replaced or eliminated if the error cannot be corrected. However, it may not be possible to determine the cause of large deviations; therefore, under practical conditions an additional sample would have to be taken if large unexplained deviations occurred.

The expected distribution of the deviations under the conditions of point two will have a mean equal to zero and a variance equal to twice the within strata variance. On this basis 99 percent of the deviations are expected to lie within the interval, $\pm 2.575 \sqrt{2\sigma_w^2}$, where σ_w^2 is the within strata variance. Values for various sampling schemes were calculated using estimates of σ_w^2 from Experiments I and II (Table 34). The distribution of

TABLE 34

99 PERCENT CONFIDENCE LIMITS OF THE DIFFERENCE BETWEEN TWO RANDOM MILK SAMPLES

Milk Constituent	Experiment Number	Number of Strata			
		None	Two	Three	Four
% Fat	I	.426	.374	.351	.351
	II	.420	.394	.393	.381
% Protein	I	.323	.293	.283	.266
	II	.259	.225	.195	.201
% Lactose	I	.270	.249	.243	.237
	II	.253	.246	.237	.232

differences under point three will have the same variance as above (point two) but will have a mean equal to the difference between two adjacent strata.

The general principle in any system designed to monitor a random sampling scheme is to select a critical value (of milk sample to sample difference) small enough so that technical errors can be detected but large enough so that deviations due to chance are ignored. Reasonable critical values for various sampling schemes are the 99 percent confidence limits shown in Table 34. Use of these values would mean that one sample in a hundred would be expected to be replaced due to chance of sampling alone if differences between strata means were zero. If a deviation exceeds the critical value the observation should be checked to determine if a technical error has occurred and if an error is detected and cannot be corrected an additional milk sample should be drawn for this strata. If no technical error is detected then the decision to retain or replace the observation will have to be based on whether the magnitude of the deviation is reasonable when the average difference between the two strata (across herds) are considered or when changes in the management or composition of the particular herd are considered. The latter part of the evaluation of deviations is somewhat subjective; however, the proposed monitoring scheme should reduce the number of observations with real errors and therefore increase the producers' confidence in the random sampling scheme. The monitoring system could be made more objective if estimates of the expected difference between strata means could be associated with the month or season in which the strata fell.

Data collected in a random sampling program should be analysed regularly so that the program can be evaluated and modified if necessary. Computer handling of milk test results make regular analyses relatively simple. Factors that should be considered in such analyses include:

1. the effect of seasonal changes in milk constituent percentages on the differences between strata means;
2. the effect of season on within herd-period variance of milk constituent percentages;
3. the identification of herds with large shipment to shipment variation in milk constituent percentages; and
4. the estimation of testing variance by regular replicate testing of milk samples from randomly selected herds.

The results of these analyses could be used to modify the sampling program for certain herds or seasons. The results could also be used to adjust the critical values (Table 34) if needed.

CONCLUSIONS

Variances associated with the procedures of sampling bulk milk and of forming composites were concluded to be small relative to the total within herd-period variances of milk constituent percentages. Sampling and laboratory (including testing) procedures used in this study--except for the formation of composites in Experiment III--were those usually followed in British Columbia and the work was done by the people who are regularly employed to do this work. Therefore, estimates of variances associated with sampling and laboratory procedures were estimates of variability under normal field conditions.

Variances associated with the laboratory analyses of milk samples were concluded to be relatively large. Milk testing procedures were found to be the main source of variation of estimates of percent lactose. Therefore the variance of estimates of percent lactose depends mainly on the number of samples analysed to obtain these estimates. Testing variances for percent milk fat and percent protein were concluded to vary from time to time.

If testing variances vary then statistically valid predictions of the variance of estimates of percent milk fat and protein cannot be made. However practical consideration do not require precise determinations of the expected variance of these estimates; a reasonable approximation is sufficient. Estimates of testing variances from; the current study, the study by Dunn [6] and the review by Green [10] indicated that the testing variances of percent milk fat and protein could usually be expected to be less than 0.007.

Biological variance accounted for approximately half the total within herd-period variance of percent milk fat and protein. Biological variance is sampling variance in the statistical sense and is, therefore, not a component of the variance of estimates of herd-period mean milk constituent percentages obtained from sampling schemes in which all shipments are sampled. Biological variance was estimated to be smaller in short periods (strata) than in long periods, but the average reduction was relatively small. However, it was concluded that stratified random sampling was worthwhile as it would be expected to reduce the frequency of large deviations from the true mean.

Significant relationships were found between within herd-period variances and milk shipment weight, percent milk fat, percent protein and percent lactose by simple and multiple linear regression techniques. However, the proportion of the sums of squares accounted for by the regression equations was relatively small for all equations. Therefore, the relationship are not useful for predicting herd-period variances.

Two-week composite samples were concluded to yield biased estimates of true means. Random samples are expected to yield unbiased estimates. Deviations of random sample estimates from the true mean should cancel out and, therefore, the mean deviation over a period of time should be close to zero.

The variance of the estimates of the mean herd-period milk constituent percentages obtained from milk samples from four randomly selected shipment was predicted to approximate the variance of estimates obtained by the compositing method currently in use. The costs associated with the collection and analyses of four randomly chosen milk samples are expected to be lower than the costs associated with the composite method now used.

Therefore on the basis of cost comparisons, expected precision and unbiasedness random sampling is to be preferred to composite sampling. The precision of estimates obtained by stratified random sampling--four strata and one sample per strata--was concluded to be acceptable,on the average, to the industry. However, for certain herds or periods the sampling frequency may need to be greater to achieve an acceptable level of precision

Alternatively costs could be reduced by taking fewer than four samples for certain herds or periods and still achieve an acceptable level of precision. The results in this study indicated that in the initial stages of a random sampling program four samples should be taken for each herd-period. The program could be assessed and modified, if necessary, by using the results obtained in the initial period. Starting a random sampling program during periods when within herd-period variance of milk constituent percentages is expected to be low (winter in this study) would reduce the probability of obtaining samples with large deviations from the true mean and allow time to accumulate data to assess the program prior to the advent of more variable seasons.

LITERATURE CITED

1. Anderson and Bancroft. 1952, Statistical Theory in Research. McGraw-Hill Book Co., New York.

2. Biggs, D.A. 1967. Milk Analysis with the Infrared Milk Analyzer. J. Dairy Sci., 50: 799-803.

3. Boswell, R.C., E. Green and D.I. Jenkins. 1967. Daily variation in the compositional quality of ex-farm milk. Brit. Milk Marketing Board. Tech. Div. Report #58.

4. Cochran, W.G. 1946. Relative accuracy of systematic and stratified random samples for a certain class of populations. Ann. Math. Stat. 17: 164-177.

5. Dimick, P.S. and H.V. Atherton. 1962. Factors influencing butterfat sampling accuracy in bulk cooled milk. Vermont Agr. Expt. Sta. Bull. 626.

6. Dunn, L.K. 1973. (unpublished data).

7. Edwards, R.A. and E. Donaldson. 1966. A study of the variability of the composition of mixed herd milks. J. Soc. Dairy Technol. 19: 110-113.

8. Fisher, R.A. and F. Yates. 1957. Statistical Tables. Oliver and Boyd, Edinburgh. 5th ed.

9. Freese, F. 1964. Linear Regression Methods for Forest Research. U.S. Forest Serv. Research Pub. FBL17.

10. Green, E. 1970. Automatic measurement of the fat and protein contents of milk. Jour. Soc. Dairy Technol. 23: 190-193.

11. Herrmann, L.F. and E.D. Anderson. 1965. Butterfat sampling and testing problems. U.S.D.A. Tech. Bull. # 1336.

12. Johnson, K.R., D.L. Fourt, R.A. Hibbs and R.H. Ross. 1961. Effect of some environmental factors on the milk fat and solids-not-fat content of cows milk. J. Dairy Sci. 44: 658.

13. Liska, B.J. and H.E. Calbert. 1954. Study of the influence of agitation time on the Babcock test of milk samples from farm bulk holding tanks. J. Milk Food Technol. 17: 14-17.

14. Morris, H.A., S.T. Coulter and C.E. Gates. 1968. Variation within herds in composition of herd milk. J. Dairy Sci. 51: 1207-1209.

15. O'Keefe, M.G. 1967. Factors affecting the design of milk total solids testing schemes. J. Dairy Res. 34: 207-210.

16. ________. 1968. The use of single or composite milk samples for the determination of fat. J. Dairy Res. 35: 291-294.

17. Preston, H.J. 1954. Developing butterfat sampling and testing programs. U.S. Dept. Agr. Farmer Co-op. Serv. Bull. 5 (52pp).

18. Snedecor, G.W. and W.G. Cochran. 1967. Statistical Methods. 6th ed. Iowa State University Press. Ames, Iowa.

19. Waite, R., J.C.D. White and A. Robertson. 1956. Variation in the chemical composition of milk with particular reference to solids-not-fat. I. The effect of stage of lactation, season of year and age of cow. J. Dairy Res. 23: 65-81.

20. Welch, B.L. 1956. On linear combinations of several variances. J. Amer. Stat. Assoc. 51: 132-148.