



100  
745  
THS



This is to certify that the

dissertation entitled

Effect of Equal and Unequal Sampling  
Intervals on Accuracy of Estimating  
Total Lactation Yield

presented by

Saloma-Lee Mildred Anderson

has been accepted towards fulfillment  
of the requirements for

Ph.D. degree in Biometry

  
Major professor

Date November 5, 1987



RETURNING MATERIALS:

Place in book drop to  
remove this checkout from  
your record. FINES will  
be charged if book is  
returned after the date  
stamped below.

|  |  |  |
|--|--|--|
|  |  |  |
|--|--|--|

EFFECT OF EQUAL AND UNEQUAL SAMPLING INTERVALS  
ON ACCURACY OF ESTIMATING TOTAL LACTATION YIELD

By

Saloma-Lee Mildred Anderson

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Department of Animal Science

1987



## **ABSTRACT**

### **EFFECT OF EQUAL AND UNEQUAL SAMPLING INTERVALS ON ACCURACY OF ESTIMATING TOTAL LACTATION YIELD**

By

Saloma-Lee Mildred Anderson

Daily milk records of 255 cows in seven herds were sampled using different frequency and spacing of samples to investigate accuracy and precision of estimating total yield. There were four sampling methods with equal intervals and six methods with unequal intervals. For schemes with unequal intervals, samples during the period of peak yield were more frequent than in other periods. For each method, total lactation yield was estimated using a linear and a nonlinear procedure. The differences between actual and estimated yield were analyzed by a fixed linear model. Interactions of sampling method with two of the fixed factors were significant. Most of the methods overestimate actual yield. Methods with fewer samples beyond day 90 generated greater biases and exhibited less precision than those with more samples beyond day 90.

## ACKNOWLEDGMENTS

I would like to thank my major professor John L. Gill, Ph. D. for his encouragement and support throughout this project. I am also grateful to the other members of my committee: Ivan L. Mao, Ph. D. for the use of the data; James H. Stapleton, Ph. D. for statistical techniques learned in his classes; and Roy S. Emery, Ph. D. for his co-operation. I am indebted to the Department of Animal Science for their support both in terms of finances and expertise upon which I relied.

Finally, I wish to express deep appreciation to my husband, Tom; my three children, Kirstin, Inger, and Erik; and my mother, Dorothea E. Shields for all they did so that I could spend all of those hours away from them in order to complete this degree. Without their understanding, I would not have continued.

## TABLE OF CONTENTS

|                                                    | page |
|----------------------------------------------------|------|
| LIST OF TABLES                                     | v    |
| LIST OF FIGURES                                    | vi   |
| 1. INTRODUCTION                                    | 1    |
| 2. REVIEW OF LITERATURE                            | 4    |
| 2.1 Introduction                                   | 4    |
| 2.2 Lactation curves                               | 5    |
| 2.2.1 Description of the curve                     | 6    |
| 2.2.2 Fitting the curve                            | 7    |
| 2.2.3 Sources of variation                         | 9    |
| 2.2.3.1 Stages of lactation                        | 9    |
| 2.2.3.1.1 Differences within                       | 9    |
| 2.2.3.1.2 Differences between                      | 11   |
| 2.2.3.2 Number and location of points on the curve | 14   |
| 2.2.3.3 Environmental factors                      | 15   |
| 2.3 Sampling methods                               | 17   |
| 2.3.1 Equally spaced methods                       | 18   |
| 2.3.2 Unequally spaced methods                     | 19   |
| 2.4 Estimation procedures                          | 20   |
| 2.4.1 Linear                                       | 20   |
| 2.4.2 Nonlinear                                    | 21   |
| 3. METHODS, RESULTS, AND DISCUSSION                | 24   |
| 3.1 Introduction                                   | 24   |
| 3.2 Data and Methods                               | 25   |
| 3.2.1 Data                                         | 25   |
| 3.2.2 Sampling methods                             | 27   |
| 3.2.3 Estimation                                   | 28   |
| 3.2.4 Models for ANOVA                             | 30   |
| 3.2.5 Models for MANOVA                            | 33   |
| 3.2.6 Model for regression                         | 35   |
| 3.3 Results and discussion                         | 36   |
| 3.3.1 Comparisons from ANOVA                       | 36   |
| 3.3.2 Biases in methods                            | 37   |
| 3.3.3 Precision of methods                         | 39   |
| 3.3.4 Yield regressed on parameters                | 44   |
| 3.3.5 Comparisons from MANOVA                      | 46   |
| 3.3.6 Biases in parameter estimates                | 48   |
| 3.3.7 Precision of parameter estimates             | 51   |
| 3.4 Conclusions                                    | 52   |
| 4. SUMMARY                                         | 54   |
| 5. BIBLIOGRAPHY                                    | 57   |

## LIST OF TABLES

|          |                                                                                                                     |            |
|----------|---------------------------------------------------------------------------------------------------------------------|------------|
| Table 1. | Frequency tabulation for classification factors.                                                                    | page<br>26 |
| Table 2. | Frequencies for ten sampling methods.                                                                               | 27         |
| Table 3. | ANOVA comparisons for deviations from actual yield for the reduced model.                                           | 38         |
| Table 4. | Comparisons of root mean squares for ten sampling methods.                                                          | 44         |
| Table 5. | Results of regressing deviations of estimated yield from actual on deviations of estimators from parameters.        | 45         |
| Table 6. | MANOVA comparisons for deviations from parameters.                                                                  | 47         |
| Table 7. | Methods that produce parameter deviations <u>NOT</u> significantly different from zero, by season.                  | 49         |
| Table 8. | Methods that produce parameter deviations <u>NOT</u> significantly different from zero, by production-parity group. | 50         |
| Table 9  | Comparisons of root mean squares of parameter deviations for ten sampling methods.                                  | 51         |

## LIST OF FIGURES

|           |                                                            |            |
|-----------|------------------------------------------------------------|------------|
| Figure 1. | Estimation of test-interval yield by test-interval method. | page<br>13 |
| Figure 2. | Biases in total yield for season.                          | 41         |
| Figure 3. | Biases in total yield for production-parity group.         | 43         |

## 1. INTRODUCTION

A statistician is asked to make inferences based on characteristics (parameters) of a population which are unknown, usually because the entire population is too large, or too widely distributed, to be measured. A subset of the population is chosen which contains most of the characteristics of the larger set and which can be easily obtained. From this subset (sample), one estimates the parameters whose values can be used with confidence in their ability to imitate the whole set. In the same manner, interval sampling of daily milk production can produce estimates of total yield.

For many years, the dairy industry has estimated yield based on equally spaced sampling methods. The intervals investigated have been tri-monthly, bi-monthly, monthly, bi-weekly, and weekly. Research over the last 20 years, suggests that sampling once every 30 days over a 305 day lactation provides estimates of milk yield whose accuracy and precision are acceptable for most purposes.

The lactation curve is not linear in time. Average daily production increases to a peak usually before day 90,

then gradually declines. Variances increase and decrease with daily averages, and greatest fluctuation occurs during maximum production.

If one samples once every 30 days, it is possible to have only two samples before day 90, which seems inadvisable vis-a-vis the characteristics of the lactation curve during this time. It is more appropriate to devise a method for sampling a cow's lactation whose intervals are unequally spaced so that most observations are collected during the time of maximum production and the remainder are collected after peak when the curve declines at a relatively constant rate.

Certain characteristics of a cow influence the shape of her lactation curve. For instance, heifers reach peak production earlier with fewer kilograms of milk than fourth parity cows. In addition to sampling method, these characteristics can be included as fixed factors in a linear model. The response variable is deviation of the estimated yield from actual yield. Using analysis of variance and the appropriate F statistics one can determine the effect of each fixed effect on the estimate of yield.

A nonlinear equation proposed by Wood (1967) contains three parameters which can be associated with characteristics of the lactation curve. The accuracy of estimates of these parameters and their effect on the accuracy of estimating total yield can be analyzed using a

regression model.

Finally, using the same parameter estimates from Wood's equation as response variables and the fixed factors as previously, one can determine whether one sampling method is best suited for all cows, or whether certain characteristics of a lactation dictate different methods.



## 2. REVIEW OF LITERATURE

### 2.1 Introduction

Total milk production from a single lactation is an important statistic for the dairy industry. Multiple records are used to measure genetic improvement and to evaluate feeding and herd management techniques. It is not always possible, nor is it economically feasible to measure each milking during lactation. However, measuring less frequently than daily introduces inaccuracy in estimation of actual yield. The introductory section of this review is about desirable characteristics of the estimator. In the next section, various aspects of the lactation curve are considered. Finally, sampling methods and procedures for estimating lactation yield are reviewed.

The need for accurate estimators in selection programs for sires and dams of future generations was emphasized by Everett, McDaniel, and Carter (1968). Further documentation was given by Menchaca (1981) whose research evaluated production of dairy herds in Cuba.

A second characteristic of a good estimator is that it should be precise, or have minimum variance. Everett,

McDaniel, and Carter (1968) compared actual yield to estimates obtained from monthly, bi-monthly, and tri-monthly sampling. They found that estimators based on more frequent sampling decreased variances of deviations from actual yield.

Estimates are obtained by sampling a cow's production. More sampling generally means higher cost. Cunningham and Vial (1968) stressed how important it is to minimize observations in the field of progeny testing, where large numbers of records are involved in estimation. McDaniel (1969) concurred by stating ". . . earlier summaries on bulls in a.i. (artificial insemination) programs [will] increase genetic progress by shortening the generation interval." One of the objectives of Menchaca's (1981) study was to minimize the number of samples.

## 2.2 Lactation curves

Milk production is curvilinear in time. A general description can be given for this relationship, but each cow has a curve with its own individual shape. Errors in estimating total yield occur because of the difficulty an investigator has in fitting a smooth mathematical curve to these individual shapes. Further sources of sampling variation are i) stage of lactation; ii) number and location of points on the curve; and iii) environmental factors.

### 2.2.1 Description of the curve

Wood (1967) described the lactation curve as beginning on day one with a value greater than zero and rising sharply to a peak, usually before day 90. It decreases much more gradually than its ascent and ends around day 305. The nonlinear equation used by Wood to describe the lactation curve is also known as an incomplete gamma function. The equation is given as:

$$y_t = at^be^{-ct} +$$

[1]

where  $y_t$  is yield at time  $t$ ,  $e$  is the base of the natural logarithm;  $a$ ,  $b$ , and  $c$  are parameters to be estimated; and  $\epsilon_t$  is the random error term.

A key factor in describing any lactation curve is establishing the time of maximum production, also known as the peak of production. In an interview with Edward Call, Switzky (1985) reported that total amount of milk produced by a cow is directly related to the amount of milk she produces at peak production. After calculating "summit milk yield" by his formula, Call demonstrates the value of this statistic for evaluating management practices.

From the literature it is clear that not all investigators agree about when peak production occurs. Kellogg et al (1977) have described peak as occurring

between days 30 and 90. Shook et al (1980) used day 40 as an average day of peak. Ferris (1981) indicated that peak production for first lactation cows usually occurs 4-8 weeks after freshening. However, Cobby and LeDu (1978) have found instances where peak occurs before the 30th day.

### 2.2.2 Fitting the curve

Although Wood's incomplete gamma function [1] generally fits a lactation, it is not the only function used by investigators. In his Doctoral dissertation, Ferris (1981) thoroughly reviewed other functions. He summarized as follows: i) Wood's equation accounted for more variation in monthly yield than an equation attributed to Nelder (1966); ii) there is evidence (Kellogg, et al; 1977) that the use of Marquardt's (1963) algorithm to estimate the parameters in Wood's equation may eliminate the need to transform data to achieve homogeneous variance; iii) examination of the residuals indicated that Wood's equation provides a better fit than other functions; iv) in general, Wood's equation provides larger  $R^2$  values.

Investigators have used two procedures to estimate the parameters in Wood's equation: i) ordinary least squares estimation after a logarithmic transformation of the equation, and ii) nonlinear least squares estimation. In his papers, Wood (1969, 1970, 1972 and 1976) estimated the parameters a, b, and c after a transformation of the data to

natural logarithms by linear least squares estimation. Cobby and LeDu (1978) compared Wood's procedure to nonlinear least squares estimation. By examining residuals they found a better fitting curve using nonlinear estimation. Lack of fit was greatest near peak with overestimation in weeks 2-10 and underestimation in weeks 11-23. They also reported an average decrease in residual mean square of 14% by using the raw data instead of log data. Congleton and Everett (1980a), also used a log transformation in estimating the parameters of Wood's equation.

Further evidence of the advisability of not using transformed data can be found when parameter estimates, thus obtained, fall outside the allowable parameter space. For example, Wood (1967) has stated that peak yield occurs when time equals  $b/c$ . A negative estimate for  $b$  or  $c$  would imply the impossible, that peak production occurred before freshening. Congleton and Everett (1980a) obtained negative estimates for the  $b$  parameter using transformed data almost three times as often when first test day occurred after the 29th day of lactation as they obtained when first test day occurred before the 10th day after freshening.

Anderson (1981) used the egg production curve of Tribolium beetles to model the lactation curve of cows. He compared the goodness of fit of four nonlinear functions to estimate egg production. After choosing an inverse cubic polynomial equation as the model for the lactation curve for

dairy cattle, he compared estimates of parameters and their standard errors obtained from autoregressive and from ordinary least squares methods. He concluded that the autoregressive method of estimation improved the accuracy of error variance, but not the accuracy of the estimates of the parameters themselves.

### 2.2.3 Sources of variation

#### 2.2.3.1 Stages of lactation

Ferris (1981) and Schaeffer et al (1977) divided lactation into three stages: before peak production, when the curve is rising; during peak production, when the curve resembles an inverted truncated parabola; and finally the period after peak, when the slope is negative as production declines. As noted by Kellogg et al (1977), each cow has a uniquely shaped curve, but most have the three distinct stages.

##### 2.2.3.1.1 Differences within stages

Of the three stages of lactation, the interval around peak is probably the most carefully studied area. It is the time when highest daily production is achieved and it is also a time when greatest variation is observed. Congleton and Everett (1980a) demonstrated these facts when their results "... showed some increase in root mean square during

the period of peak production." Similarly, Cobby and LeDu (1978) used Wood's equation [1], but they compared its fit to two nonlinear alternative equations. They also found that lack of fit is greatest near peak production, even though they were using transformed data which should give less weight to high yields. They concluded that there are major problems in fitting lactation curves near the time of peak production because so few samples are taken prior to this time that it is difficult to estimate response in this region or any parameter associated with it.

In the initial stage of lactation when the curve is rising at a rapid rate, Congleton and Everett (1980a) found that bias and error can be high as early as the first week. They concluded that length of interval to first test day influenced the shape of the curve. Everett et al (1968) reported that 80% of the bias occurred when estimating yield in the first month of lactation. Cobby and LeDu (1978) recommended that investigators sample more frequently early in lactation. In particular, if the first test day falls on day 30 of lactation, so much information will be lost that there will be little reliability in the final estimate of yield.

The interval from peak to near day 290 of lactation exhibits less variable behavior than the days prior to this time. After studying many lactations, Cunningham and Vial (1968) suggested that researchers can "cease recording after





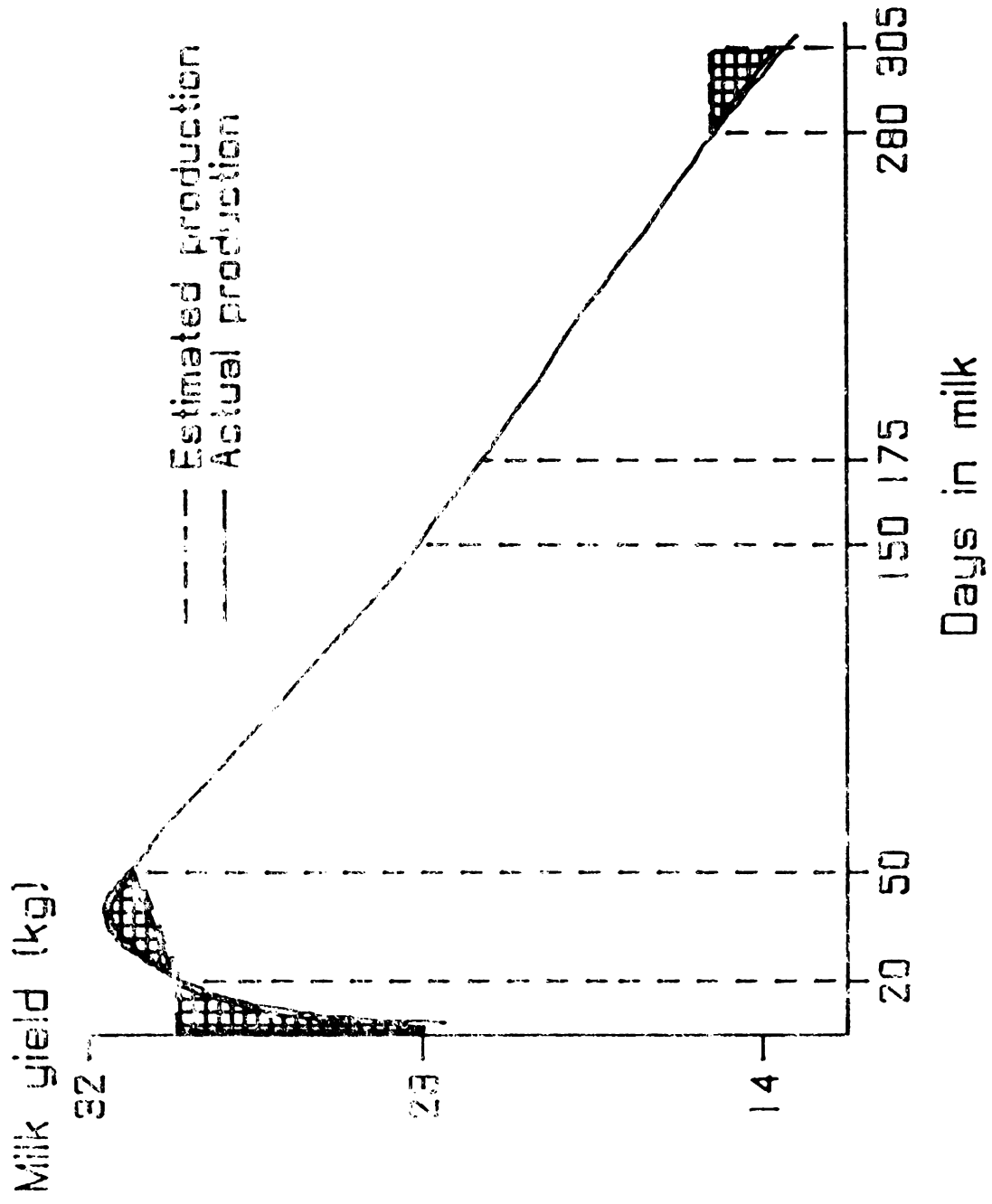
the first few months and record. . . dry-off date. The terminal portion of the curve appeared to decline linearly." This consistent behavior of the portion of the lactation curve from peak to near dry-off date was further emphasized by Shook et al (1980), because they did not provide adjustment factors for this period.

Finally, for those cows whose lactation continues until day 305, Schaeffer et al (1977) stated that the variability found in early lactation is again present when a test day falls after 290 days in milk.

#### 2.2.3.1.2 Differences between stages

Because lactation is not constant over time, allowances must be made for estimates obtained at different stages of lactation. McDaniel (1969) reviewed sixty reports to compare estimation of lactation yields with samples taken at various intervals. He emphasized that adjustments must be made to an estimate of total yield based on the stage of lactation in which samples are collected. Shook et al (1980) used Figure 1 to graphically describe the relationship between total yield estimates based on samples from particular stages of lactation. They demonstrated three instances when the estimates are biased because of the shape of the curve: 1) the interval from calving to first test overestimates actual yield; 2) the interval around lactation peak underestimates and 3) the interval from last sample to

**Figure 1.** Estimation of lactation yield by test interval method. Area under lactation curve represents actual yield; areas enclosed by dashed lines represent estimated yield; and cross-hatched areas represent amount by which the estimate is biased.





dry-off day, again, overestimates actual yield. The purpose of the article was to provide the adjustment factors to which McDaniel (1969) had alluded. Kellogg et al (1977) further emphasized the differences between stages of lactation by noting that individual cows vary more in their second month than they do in their eighth month of lactation. Instead of dividing lactation into three stages, Schaeffer and Burnside (1976) hoped to obtain better estimates by using 16 distinct stages.

#### 2.2.3.2 Number and location of points on the curve

O'Connor and Lipton (1960) used intervals of 7, 14, 28, 42, 56 and 63 days and deviated estimated from actual milk yield. Their estimation procedure was a linear function of first and last sample day and of interval length. They found that mean differences and errors of estimation increase as the length of the sampling interval increases. McDaniel's (1969) review included comparisons of monthly, bi-monthly, and tri-monthly sampling to actual yield. He concluded that ". . . average error in lactation is primarily a function of the length of the interval between tests." Cunningham and Vial (1968) compared monthly deviations from actual to two kinds of bi-monthly testing schedules, one beginning between days 4 and 34, the other between days 34 and 64. Like O'Connor and Lipton (1960) and McDaniel (1969), they also found smaller mean squared

errors for the shorter (monthly) interval.

Barta and Lee (1985) used two linear and one nonlinear procedures to estimate total yield for cows of four different breeds and two different age groups. The magnitude of the biases of the nonlinear results was less than the two linear ones while the standard errors of prediction were similar for all breeds and both age groups. The nonlinear procedure was less biased and more accurate than the other two methods in mid-lactation. This is of interest to those who use this stage of lactation for calculating estimates or making projections about actual yield. These results of Barta and Lee agree with earlier results published by Schaeffer et al (1977) when using the same three kinds of estimation procedures.

#### 2.2.3.3 Environmental factors

The effect of a cow's environment on the shape of the lactation curve, and hence the estimate of total yield, is well documented in the literature. Most investigators include these influences as fixed factors in a linear model to analyze their effect on total milk production. Wood (1969) used his equation [1] to fit weekly records which began in different months. An analysis of variance showed that month of calving (season of calving) was a significant factor when determining total yield and also when estimating the three parameters in his equation. Further documentation

of the effect of seasonality was given by Schaeffer et al (1977) who stated that the slope for the declining portion of lactation is different for different seasons. Congleton and Everett (1980b) list tables by month of calving. The tables contain the estimates of the three parameters (a, b and c) in Wood's equation. The effect of seasonality is shown when a produces largest values in early summer, but b and c peak in winter. Season of calving is also listed as an important fixed factor by Schaeffer and Burnside (1976), Keown et al (1986), Barta and Lee (1985), Miller et al (1970), Menchaca (1981) and in a later paper by Wood (1976).

Generally speaking, a herdsman feeds and otherwise tends to cows in his herd in a particular way. Therefore, differences between herds should be more pronounced than differences within herds. Schaeffer and Burnside (1976) and Keown et al (1986) have suggested that differences in herd management practices can be accounted for by including this factor in a model. On the other hand, Wood (1970) found little change in the shape of the curve due to the variation in management, and he suggested that it need not be a factor in a model.

A cow's parity affects the amount of milk given during a lactation and the need for its inclusion in a model is documented by Wood (1969), Congleton and Everett (1980b), Keown et al (1986), Miller et al and Schaeffer and Burnside (1976). In addition, Wood (1969) suggested that although





first, second, and third lactations can be different for one cow, it is unlikely that the fourth or higher lactations will vary much. He advocated the combination of third or higher lactations to form subclasses for a parity factor in a model. When the past production level of a cow was known, Congleton and Everett (1980b) included this factor in a model.

Certain continuous variables are also associated with the amount of milk given by a cow. Those can be included as covariates in a linear model. It is known that the more days a cow remains open (unbred) during a lactation translates to more milk produced than if she were bred earlier. Schaeffer and Burnside (1976), Keown et al (1986), Schaeffer et al (1977) and Miller et al (1967) included days open in their models. It is also known that the more days a cow is milked the higher her total production will be, so Cunningham and Vial (1968) use "days in milk" as a covariate in their model.

### 2.3 Sampling methods

The methods for sampling a cow's milk production can be divided into two categories: equal intervals between sample days (the method most often used); and unequal intervals between sample days, which is the focus of this dissertation.



### 2.3.1 Equally spaced sampling intervals

In McDaniel's (1969) review of monthly, bi-monthly, and tri-monthly sampling methods he concluded that when evaluating herd improvement programs "...monthly testing... is accurate enough for practical management...". Sargent et al (1968) compared monthly and bi-monthly records of Holstein and Guernsey herds and found that both methods were biased but not appreciably different from each other. Everett, McDaniel, and Carter (1968) used monthly, bi-monthly, and tri-monthly collection schemes to compare adjusted and unadjusted estimates of yield, and used the method of the Dairy Herd Improvement Association (DHIA) in use at that time. They concluded that the use of adjustment factors for the beginning and end of lactation yielded more accurate estimates of total yield, and that biases in estimates were proportional to the length of the intervals between the days. In a previous article, Everett, Carter, and Burke (1968) used monthly or less frequent sampling and compared various estimation procedures. A reduction in bias was accomplished by adjusting estimates "according to the day of lactation on which the test occurs." They also stressed that a majority of the bias occurs in the first test period. Menchaca (1981) used weekly, biweekly, and monthly sampling methods with three different starting days and three different estimation procedures. He observed that biases increase as the length of the sampling interval

increases, and that estimation procedures and sampling methods should be taken into account when considering the accuracy of the estimate of total yield. Badner et al (1984) compared lactation curves fitted from daily, weekly, bi-weekly, and monthly milk weights. Within their fixed effects model, sampling frequency had no effect on the parameter estimates from Wood's equation. O'Connor and Lipton (1960) used 18 lactations from 12 Shorthorns and sampled at each of 7, 14, 28, 42, 56, and 63 days. They also found bias in estimating milk yield that increased as the length of the intervals increased.

#### 2.3.2 Unequally spaced sampling intervals

Alexander and Yapp (1949) compared four methods of sampling a cow's yield to monthly tests by deviating estimates from actual yield. Two of these four methods contained unequally spaced intervals. One of them that sampled in the 2nd, 4th, and 10th month was accurate enough to suggest its use when cost of hand calculation must be lowered in order to increase the number of cows tested ". . . and accelerate improvement of dairy herds." Their data accounted for about 4% of the estimated 1,123,000 cows in Illinois at that time.

## 2.4 Estimation procedures

Procedures for estimating total yield may be placed in two categories: linear interpolation between sample points and estimation of parameters for nonlinear functions based on sample points.

### 2.4.1 Linear

Herd improvement programs use the centering date method or the test interval method to estimate total yield. Both procedures linearly interpolate between sample days to estimate yield for that interval. Shook et al (1980) noted that the centering date method of linear interpolation between sample points requires a rigid test day schedule. Sargent et al (1968) compared the two methods and reported that i) they appear to be equally accurate and ii) that differences between them result more from sampling error in the test-day milk weights than from differences in the two methods.

Calculation of the estimate of total yield using the test interval method described by Wiggans and Grossman (1980) proceeds as follows. A cow's lactation is divided into intervals by the sample days. Production estimates for the first half of the interval are calculated from the previous sample day information. Estimates for the second half of the interval are calculated from the present sample day's information. The two estimates are added to obtain



the "test interval credit" which is then summed over all samples. First and last intervals are given credit based on one sample and added to the total to complete the lactation estimate. Factors can be used according to Shook et al (1980) to adjust the first and last test day where the test interval procedure overestimates yield. During peak production the curve is convex and factors can again be used to compensate for underestimated yield. The proposed factors are functions of five variables: a cow's parity; the day of the first sample vis-a-vis freshening; the day of the previous sample; the length of the sampling interval; and the day of the last sample. No attempt is made to adjust for season of lactation. In their conclusions, Shook et al (1980) claimed that test interval factors contained in their tables effectively reduce bias and sampling error in estimates of test interval yield.

#### 2.4.2 Nonlinear

The equation [1] proposed by Wood (1967) and the justification for its use were discussed in Section 2.2 on lactation curves. What follows here are some results others have reported using different equations. Also discussed is the method used to obtain estimates of the parameters.

One of the objectives of the investigation conducted by Schaeffer et al (1977) was to propose a nonlinear function which described the three stages of lactation:



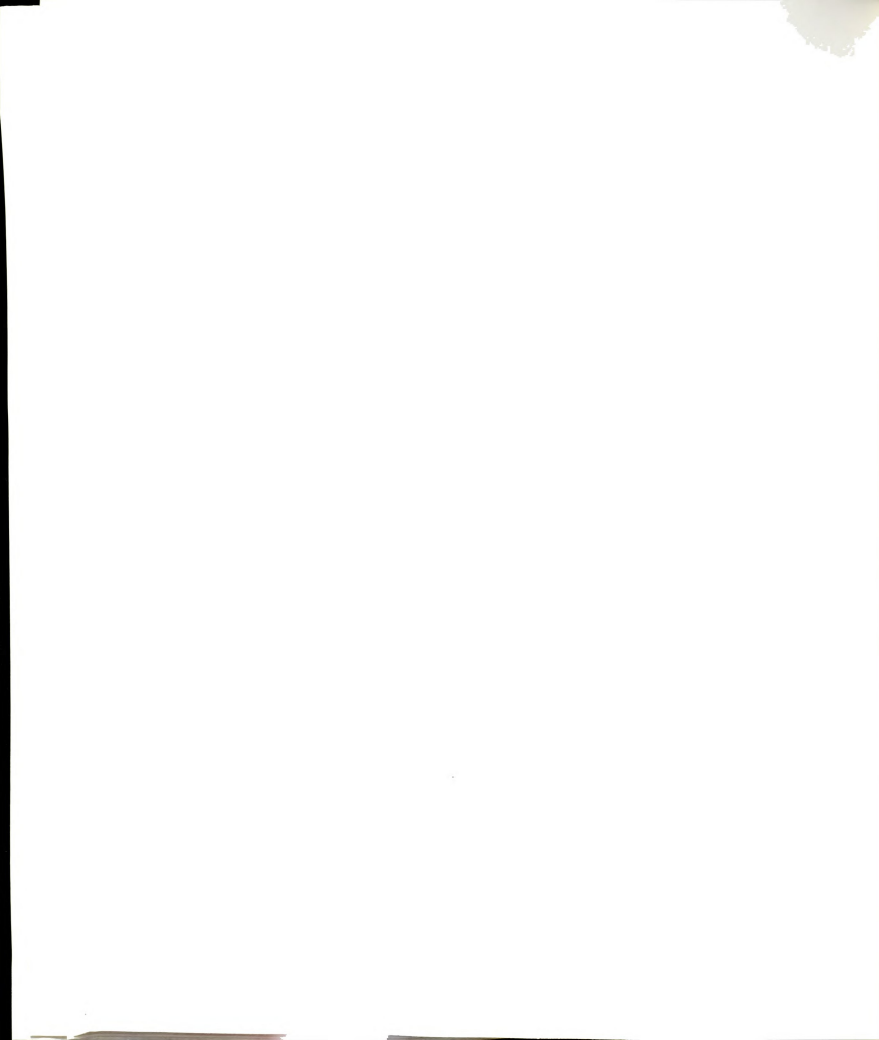


before, during, and after peak production. Their model for milk yield was a function of time, and it contained six parameters to be estimated. One of them is difficult to estimate when individual records begin as late as the 6th day after freshening, which is not unusual for data collected from the field.

Cobby and LeDu (1978) suggested two new models which are based on Wood's [1], but which incorporate the influence of "...maximum yield and persistency (defined as the extent to which peak yield is maintained)." The goodness of fit of the new models was compared to Wood's [1]. Results indicate that one of the models should not be considered and the other was on the average as accurate as Wood's.

In a similar manner, Badner and Anderson (1985) fit four nonlinear models to data from 94 lactations. One of these equations was Wood's [1], the second was a variation of Wood's equation, the third is attributed to Nelder (1966), and the fourth is based on the third. When all models were adjusted for autocorrelation that exists when samples are taken on the same animal, the fit of all four of the models was essentially equal.

All nonlinear equations used to estimate milk yield contain parameters that must be estimated. Marquardt (1963) presented an algorithm to be used in the least-squares estimation of nonlinear parameters. It is a compromise between two previously used methods ( steepest descent and



Gauss) which iteratively minimizes the sum of squares for error based on the direction of steepest descent or the distance to the next lowest point. Marquardt's algorithm is also thought to be the most appropriate one to use when there is reason to believe that the parameters are correlated.

Finally, Congleton and Everett (1980b) compared Wood's equation to the test interval method of obtaining estimates of milk yield. They concluded that the incomplete gamma curve can be used with comparable or higher accuracy than the test interval method, despite the fact that "Shook factors" are used to increase accuracy of the test interval method.

### 3. METHODS, RESULTS, AND DISCUSSION

#### 3.1 Introduction

Daily measurement of milk production for an entire lactation of a cow provides the most accurate measure of a cow's total yield. Less expense and time is required if a cow is sampled less frequently, yet often enough to support reasonably accurate estimation. In the past, most sampling has been performed using intervals of equal length. McDaniel (1969) reviewed 60 research reports dealing with the accuracy of estimating lactation yield from samples taken at monthly, bimonthly, and trimonthly intervals. He concluded that monthly sampling produces estimates that are within 5% of actual yield and that error of estimation increases as the length of sampling intervals increases. Menchaca (1980) and Badner et al (1984) sampled less frequently than monthly obtaining similar results to McDaniel for estimation error.

The lactation curve is not linear over time. Shook et al (1980) noted that there are three periods when linearly interpolated estimates are biased: before peak production, during peak production, and the last fifteen days of

lactation. Congleton and Everett (1980) also documented the influence of sampling during peak production on accuracy when they observed larger mean squares during that period.

The objective of this investigation is to determine the effect of unequally spaced intervals, with varied degrees of emphasis at peak period, on the accuracy and precision of estimating total yield. To permit comparison with standard practice, records also were sampled at equally spaced intervals of different lengths. Also investigated was the accuracy of nonlinear estimation of various characteristics of lactation curves on the accuracy of estimating lactation yield.

### 3.2 Data and Methods

#### 3.2.1 Data

Daily milk weights for 405 cows of various ages in seven herds were obtained from Eli Lilly and Company. Records of 150 animals were discarded because lactation ended before the 200th day or because other information was absent. Some daily records were considered outliers by the criterion that production should not differ from that of the preceding day by more than 40%. Outliers and all records of zero weights were replaced by linear interpolation.

The data were subdivided into classes according to season of freshening, herd, and dietary treatment (the data

|                             |            |           |           |            |            |            |            |
|-----------------------------|------------|-----------|-----------|------------|------------|------------|------------|
| Season                      | J-F<br>39  | M-A<br>22 | M-J<br>14 | J-A<br>54  | S-O<br>76  | N-D<br>50  |            |
| Herd(season)                | J-F        | M-A       | M-J       | J-A        | S-O        | N-D        |            |
| herd 1                      | 0          | 0         | 1         | 20         | 14         | 8          | 43         |
| herd 2                      | 19         | 20        | 7         | 1          | 0          | 0          | 47         |
| herd 3                      | 2          | 0         | 2         | 10         | 21         | 11         | 46         |
| herd 4                      | 10         | 0         | 0         | 4          | 16         | 14         | 44         |
| herd 5                      | 0          | 0         | 0         | 14         | 17         | 5          | 36         |
| herd 6                      | 5          | 1         | 2         | 3          | 4          | 12         | 27         |
| herd 7                      | 3          | 1         | 2         | 2          | 4          | 0          | 12         |
| Treatment                   | 1<br>60    | 2<br>66   | 3<br>56   | 4<br>73    |            |            |            |
| Production-<br>parity group | heif<br>74 | hi2<br>18 | hi3<br>24 | med2<br>58 | med3<br>43 | low2<br>25 | low3<br>13 |
|                             |            |           |           |            |            |            | total=255  |



### 3.2.2 Sampling methods

Lactation yields were estimated from two kinds of sampling: equally and unequally spaced intervals. Equally spaced sampling methods of 30-day, bi-weekly, weekly, and 3-day intervals were investigated. Two kinds of unequally spaced sampling methods were developed. The first used ten sample points, the same number as in the 30-day equal method; the second used 22 observations, the number in the bi-weekly method. Table 2 shows the frequencies for each sampling method.

**Table 2.** Frequencies for ten sampling methods.

|                 | Days in milk |       |        |
|-----------------|--------------|-------|--------|
|                 | 0-14         | 15-90 | 91-305 |
| Sampling method |              |       |        |
| 10/01           | 1            | 8     | 1      |
| 10/02           | 2            | 7     | 1      |
| 10/03           | 1            | 6     | 3      |
| 10/04           | 1            | 5     | 4      |
| 30-day          | 1            | 2     | 7      |
| 22/01           | 2            | 15    | 5      |
| 22/02           | 2            | 10    | 10     |
| bi-weekly       | 1            | 6     | 15     |
| weekly          | 2            | 11    | 30     |
| 3-day           | 3            | 26    | 70     |





Highest daily production usually is achieved during the interval of 15 to 90 days in milk. Congleton and Everett (1980a) found that greatest variation occurred near peak production. Cobby and LeDu (1978) found problems in fitting lactation curves near the time of peak production because so few samples are taken prior to that time. Therefore, a heavier concentration of sampling in early lactation might be warranted. In these data, average day of peak yield was approximately day 40 (standard deviation of 25.6 days). A high proportion of the cows achieved peak production from day 15 to day 90. In Table 2, for methods based on 10 sample points, the first two methods (10/01 and 10/02) concentrate sampling most heavily before and during peak production. The next two methods (10/03 and 10/04) remove sample days from the interval around peak and place them after peak. Finally, method 10/05 is equally spaced with 30-day intervals and contains fewest samples before day 90. In the same table, for methods based on 22 sample points, the progression from most to least samples before day 90 is again repeated for methods 22/01, 22/02, and bi-weekly. The last two methods listed are equally spaced weekly and 3-day sampling.

### 3.2.3 Estimation

Procedures for estimating total yield were: (1) linear interpolation between sample points and (2) estimation of

parameters for nonlinear functions based on sample points. The Dairy Herd Improvement Association (DHIA) has adopted the test interval method of estimating between sample days described by Wiggans and Grossman (1980), with the early and late samples adjusted by using "Shook factors" (Shook et al, 1980). For each proposed sampling method an estimate of total yield was obtained using the test interval method and recorded as deviation from actual yield. This procedure will be referred to as the "linear" portion of the investigation.

The second estimation procedure fitted each cow's daily milk weights to a model which was not linear in parameters and was proposed by Wood (1967). It is given as:

$$y_t = at^be^{-ct} + \epsilon \quad [1]$$

where  $y_t$  is the estimate of yield at time  $t$ ;  $a$ ,  $b$ , and  $c$  are parameters to be estimated,  $e$  is the base of the natural logarithm, and  $\epsilon$  is a random error term. Marquardt's algorithm (1963) in PROC NLIN of SAS Institute Inc. (1985) was used to obtain estimates of the parameters  $a$ ,  $b$ , and  $c$  based on each of the 10 sampling methods. These estimates were then used to obtain estimates of total yield which were deviated from actual yield.



### 3.2.4 Models for ANOVA

Actual total yield was obtained by summing daily milk weights. The dependent variable for our model was the difference between actual yield and an estimate based on the linear or nonlinear procedure. The model was:

$$\begin{aligned}
 Y_{ijkmnr} = & \mu + S_i + H(S)_j(i) + T_k + P_m + \\
 & X_{1ijkmn} + X_{2ijkmn} + \text{error } l_{ijkm(n)} + \\
 & M_r + (M*S)_{ir} + (M*P)_{mr} + \text{error } 2_{ijkmnr}
 \end{aligned} \quad [2]$$

where  $Y_{ijkmnr}$  is the deviation from actual production using sampling method  $r$  on the  $n$ th cow in production-parity group  $m$ , treatment group  $k$ , and herd  $j$ , within season  $i$ ;

$\mu$  is the overall mean;

$S_i$  is the fixed effect of the  $i$ th season ( $i=1..6$ );

$H(S)_j(i)$  is the fixed effect of the  $j$ th herd nested within the  $i$ th season;

$T_k$  is the fixed effect of the  $k$ th dietary treatment ( $k=1,2,...,4$ );

$P_m$  is the fixed effect of a particular production group and parity combination ( $m=1,2,...,7$ );

$X_{1ijkmn}$  is the covariate days open;

$X_{2ijkmn}$  is the covariate days in milk;

$\text{error } l_{ijkm(n)}$  is the random effect of the  $n$ th cow within a particular combination of the aforementioned factors



( $n=0,1,2,\dots$ ) with error 1 (cow) as an independent random variable distributed as  $N(0, \sigma_1^2)$ ;  
 $M_r$  is the fixed effect of the  $r$ th sampling method ( $r=1,2,\dots,10$ );  
 $(M*S)_{ir}$  is the interaction of sampling method  $r$  and season  $i$   
 $(M*P)_{mr}$  is the interaction of sampling method  $r$  and production-parity group  $p$ ; and, finally,  
error 2<sub>ijkmnr</sub> is residual random error with error 2 (residual) as an independent random variable distributed as  $N(0, \sigma_2^2)$ , and further, error 1 and error 2 are assumed independent.

Other possible interactions were deemed unimportant, a priori.

To analyze data according to this model in traditional univariate procedures, Geisser and Greenhouse (1959) imply that one must assume (i) the expected value of the deviation from actual yield depends only on the combinations of named fixed factors to which each record belongs and (ii) that the collection of deviations estimated from several methods of sampling the same daily records of a cow, when considered as multivariate data, have a covariance matrix  $\Sigma$  that is independent of the fitted factors and has the following form:

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho & \rho & . & . & . & \rho \\ & 1 & \rho & & & & & \rho \\ & & 1 & \rho & & & & \rho \\ & & & . & & & & . \\ & & & & . & & & . \\ \text{symmetric} & & & & & . & & . \\ & & & & & & 1 & \end{bmatrix} (\sigma_1^2 + \sigma_2^2), \quad \text{for } \rho = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad [3]$$

where  $\rho$  is the correlation between observations on the same cow for any two methods of sampling. In this investigation we believed that the covariance matrix  $\Sigma$  for the 10 sampling methods would not be uniform, even if the cow variances ( $\sigma_1^2$ ) and residual variances ( $\sigma_2^2$ ) were homogeneous. For instance, methods based on 10 observations and sampled heavily before day 90 are likely to be closely correlated; whereas, these methods would not be expected to closely correlate with methods which are equally spaced every three days.

It has been suggested by Cole and Grizzle (1966) that when the assumption of uniform covariance for a repeated factor (sampling method) cannot be justified, that multivariate methods be applied to account for the inherent heterogeneous covariances. The model, rewritten to conform to multivariate ANOVA techniques, can be stated in matrix notation as:

$$Y = X \beta + \varepsilon \quad [4]$$

where  $Y$  is a (255x10) matrix whose column vectors are deviations from actual production for sampling method  $r$  ( $r=1,2,\dots,10$ ) and whose row vectors, corresponding to cows, are independently multivariate normally distributed with covariance matrix  $\Sigma^*$ , not necessarily of the form given at [3].





$X$  is a (255x51) design matrix of rank 47 corresponding to the first seven factors of the model written at [2];

$\beta$  is a (51x10) matrix of parameters; and

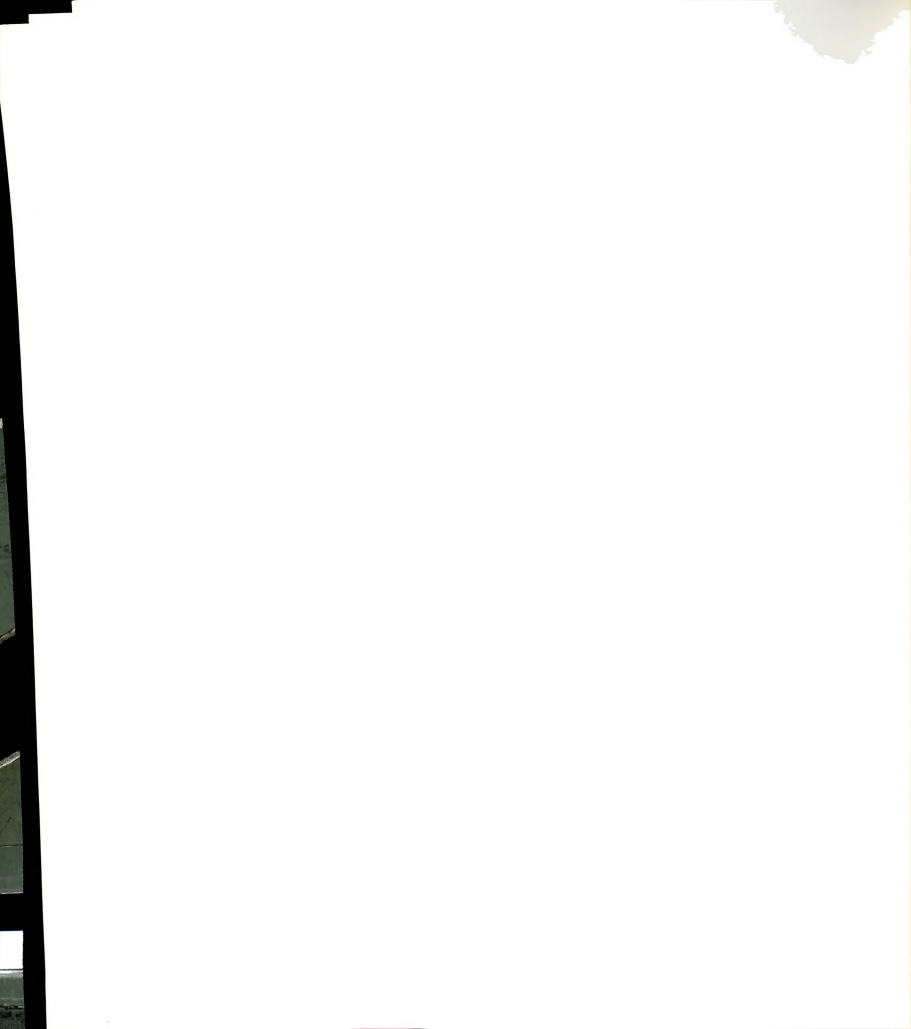
$\epsilon$  is a (255x10) matrix of residual random errors.

To analyze the linear model [4], the REPEATED option in PROC GLM of SAS Institute Inc. (1985) was used.

### 3.2.5 Models for MANOVA

Wood (1967, 1972) has given the following biological interpretation to the three parameters,  $\underline{a}$ ,  $\underline{b}$ , and  $\underline{c}$  found in his equation [1]:  $\underline{a}$  is a constant representing the amount of milk in the mammary gland at freshening;  $\underline{b}$  represents the slope of the curve to peak production; and  $\underline{c}$  represents the rate of decline after peak. Estimates ( $\tilde{a}$ ,  $\tilde{b}$ , and  $\tilde{c}$ ) of these parameters based on daily records should be close to actual values of parameters for each cow. Estimates ( $\hat{a}_r$ ,  $\hat{b}_r$ ,  $\hat{c}_r$  for  $r=1,2,\dots,10$ ) based on each of the ten sampling methods will necessarily be somewhat less accurate. Each cow, then, has three sets of deviations  $\hat{a}_r - \tilde{a}$ ,  $\hat{b}_r - \tilde{b}$ , and  $\hat{c}_r - \tilde{c}$ ; for  $r=1,2,\dots,10$  sampling methods.

To study the relationship between sampling methods and parameter  $\underline{a}$  in Wood's equation [1], an analysis of variance for repeated measures was performed using each of the estimated parametric deviations as a response variable with the same fixed factors as in the linear model [4].



The model was:

$$Y_a = X \beta_a + \epsilon_a \quad [5]$$

where  $Y_a$  is a (255x10) matrix whose columns correspond to the sampling methods and whose rows correspond to cows, so that a typical entry,  $\hat{a}_{nr} - \tilde{a}_n$ , is the difference, for cow  $n$  ( $n=1,2,\dots,255$ ), between an estimate for parameter  $a$  using sampling method  $r$  ( $r=1,2,\dots,10$ ) and the estimate for  $a$  using daily records;

$X$  is the (255x51) design matrix found in [4];

$\beta_a$  is a (51x10) matrix of parameters; and

$\epsilon_a$  is a (255x10) matrix of residual random errors.

A similar set of equations can be defined for the other two parameters,  $b$  and  $c$ , in Wood's equation to study their relationships with the ten sampling methods.

Finally, to study the relationships among the three parameters, simultaneously, and how they effect the ten sampling methods, a multivariate analysis of variance (MANOVA) for repeated measures was performed on the parametric deviations using the same fixed factors as in [4] and [5]. The MANOVA model was:

$$Y^* = X \beta^* + \epsilon^* \quad [7]$$

where  $Y^*$  is a (255x30), such that

$$Y^* = [Y_a, Y_b, Y_c]$$



$X$  is the (255x51) design matrix of [4] and [5];

$\beta^*$  is three concatenated (51x10) matrices of parameters to be estimated, one matrix for each set of dependent variables and can be written as:

$$\beta^* = \begin{bmatrix} \beta_a \\ \beta_b \\ \beta_c \end{bmatrix}$$

$\epsilon^*$  is a matrix of error terms similarly partitioned as  $Y^*$ .

### 3.2.6 Model for regression

In model [4], the dependent variable was the difference between actual yield and an estimate of yield based on a sampling method. In models [6] and [7], the dependent variable(s) were differences between estimates of parameters based on daily records and estimates of parameters based on a sampling method. The independent variables for all three models were the same fixed factors. In contrast, if one uses the parameter deviations as independent variables and yield deviations as the dependent variable, one can examine the contribution of these curve characteristics to the accuracy of each sampling method.

The regression of deviations from actual yield on estimated parameter deviations for each sampling method was performed using the following model:

$$d_r = X_r \beta_r + \epsilon_r \quad [8]$$

where  $d_r = \hat{y}_r - y$  is a (255x1) vector of deviations from actual



yield based on sampling method  $r$  (for  $r=1,2,\dots,10$ );  
 $X_r$  is a  $(255 \times 4)$  matrix whose last three columns are vectors  
of deviations  $\hat{a}_{nr} - \tilde{a}_n$ ,  $\hat{b}_{nr} - \tilde{b}_n$ ,  $\hat{c}_{nr} - \tilde{c}_n$  for sampling  
method  $r$  and cow  $n$  ( $n=1,2,\dots,255$  and  $r=1,2,\dots,10$ );  
 $\beta_r$  is a  $(4 \times 1)$  vector of regression parameters sampling  
method  $r$ ;  
 $\epsilon_r$  is a  $(255 \times 1)$  vector of random error terms for sampling  
method  $r$ .

The usual regression assumptions were made and the model was used to obtain regression parameter estimates for each of the ten sampling methods. Examination of standardized partial regression coefficients and partial  $F$  tests should determine the relative magnitude of contribution from these curve parameters to the accuracy of each sampling method.

### 3.3 Results and Discussion

#### 3.3.1 Comparisons from ANOVA

An analysis of variance (ANOVA) was performed on the linear model [4] using yield deviations as the dependent variable. Factors whose observed significance level was greater than .2 were removed from the model and the analysis was performed again. After a factor was removed from the model it was not re-entered. The process stopped when all factors remaining in the model fit the criterion. The final results for the analysis of deviations on kilograms of milk





produced for the linear and non-linear procedures are shown in Table 3. Interaction of method with season was significant ( $P < .07$ ), as was interaction of method with production-parity group ( $P < 0.04$ ). Both estimation procedures produced analysis with the same factors in the linear model at approximately the same observed significance levels.

### 3.3.2 Biases in methods

To understand bias in the estimate of total yield, it was useful to identify subclasses involved in significant interactions whose mean deviations were significantly ( $P < .05$ ) different from zero. Figure 2 graphically represents those subclasses for the interaction of season with sampling method for the two estimation procedures: linear and nonlinear. The vertical axes are average deviations of milk and the horizontal axes represent the six seasons with numbers of observations for each subclass in parentheses. The interval for each season which is darkened represents half of an acceptance region for testing the hypothesis that average deviations are zero. These intervals were calculated according to Tukey's minimum significant difference as outlined in Gill (1978). In Figure 3 production-parity group replaces season.

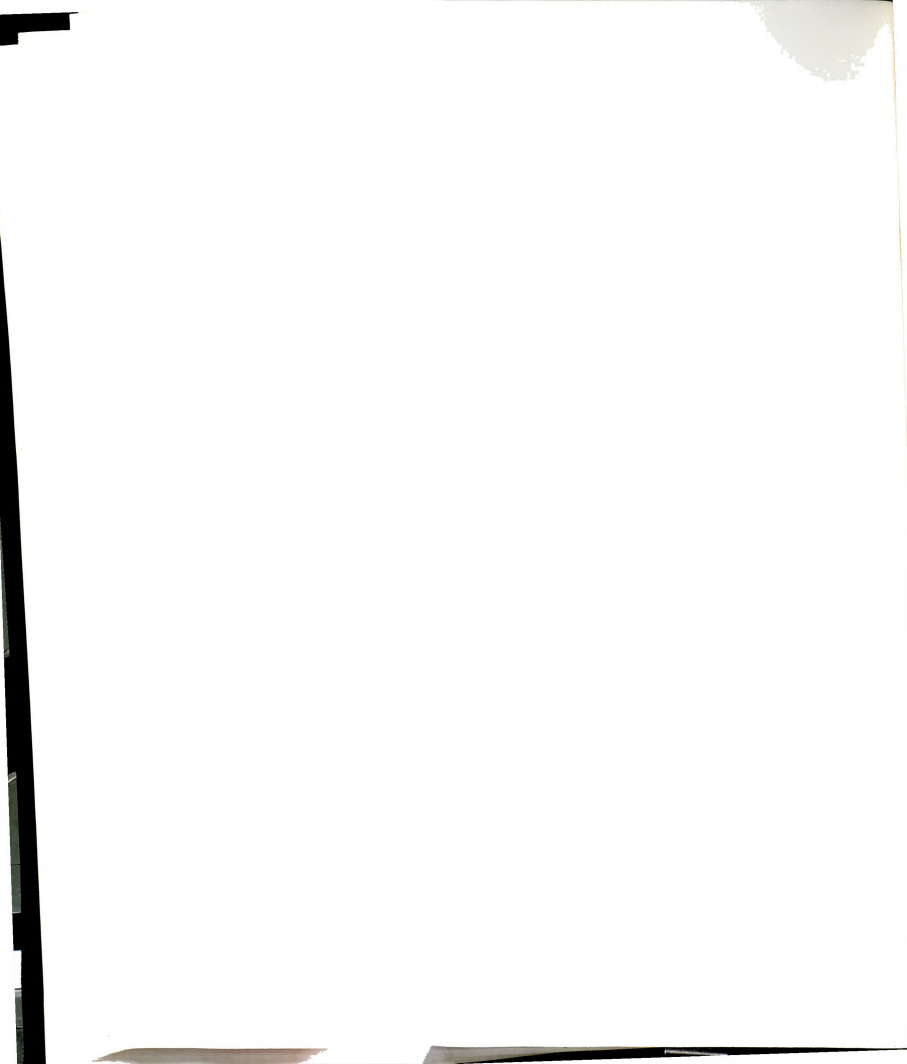


Table 3. ANOVA comparisons for deviations from actual yield for the reduced model.

| Model factor            | df   | linear estimates |       | nonlinear estimates |       |
|-------------------------|------|------------------|-------|---------------------|-------|
|                         |      | F value          | P > F | F value             | P>F   |
| season                  | 5    | 4.12             | .0013 | 5.53                | .0001 |
| herd(season)            | 25   | 1.73             | .0205 | 2.24                | .0011 |
| treatment               | 3    | NA               | NA    | NA                  | NA    |
| production-parity group | 6    | 3.26             | .0043 | 2.27                | .0381 |
| days open               | 1    | 10.08            | .0017 | 7.13                | .0082 |
| days in milk            | 1    | NA               | NA    | NA                  | NA    |
| error 1 (cow)           | 213  |                  |       |                     |       |
| method                  | 9    | 3.03             | .0020 | 2.98                | .0023 |
| method*season           | 45   | 1.35             | .0622 | 1.42                | .0390 |
| method*p-p              | 54   | 1.42             | .0258 | 1.38                | .0380 |
| error 2 (res)           | 2187 |                  |       |                     |       |

NA (not applicable): this factor was not included in the reduced model

Generally speaking, for both the linear and nonlinear procedures, biases most often occurred when samples were unequally-spaced and numbered only ten. This is true across all seasons and for all production-parity groups. For the linear procedure, no method had mean deviation significantly



different from zero in the months of September through December where half of the observations were obtained. For the production-parity groups, most significant deviations occurred in classes with the fewest observations. Methods based on 22 sample points, whether those were unequally or equally spaced, had mean deviations significantly different from zero about one-third as often as the sampling methods with ten unequally spaced points. Those occurred in classes which had the smallest number of observations. For the nonlinear procedure of estimation, the ratio of occurrence of significant bias in methods is about three to one for methods based on ten unequally-spaced points to any of those with 22 sample points. Also, weekly sampling contributes more bias in yield estimated nonlinearly than linearly.

### 3.3.3 Precision of methods

To evaluate the precision with which the sampling methods estimate total yield one can compare their variances; or, alternatively, their root mean square residual errors. Table 4 contains these statistics ranked within linear and non-linear procedures. The two highest ranking methods (the two with the most samples: weekly, 3-day) are the same for the linear and nonlinear procedures. The same is true for the two lowest ranking (the unequally-spaced points with fewest observations beyond peak lactation). For each method, the root mean square

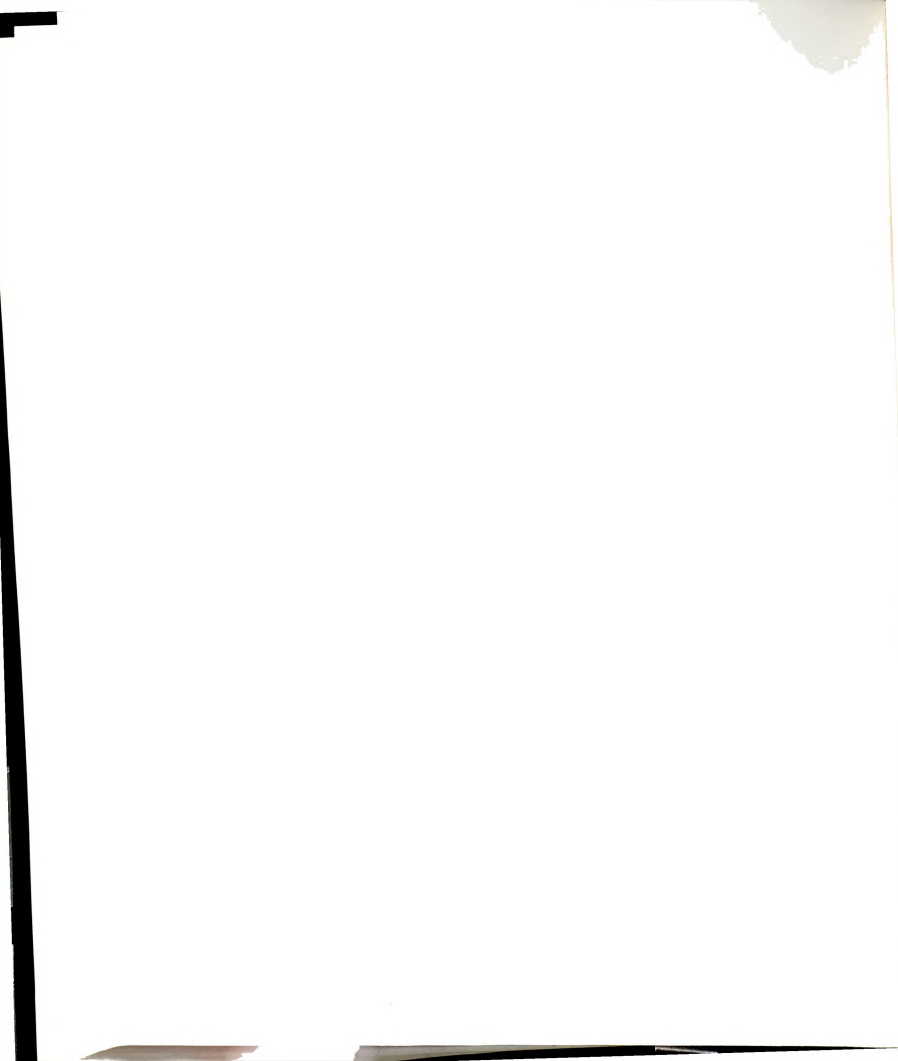


Figure 2. Biases in total yield for season





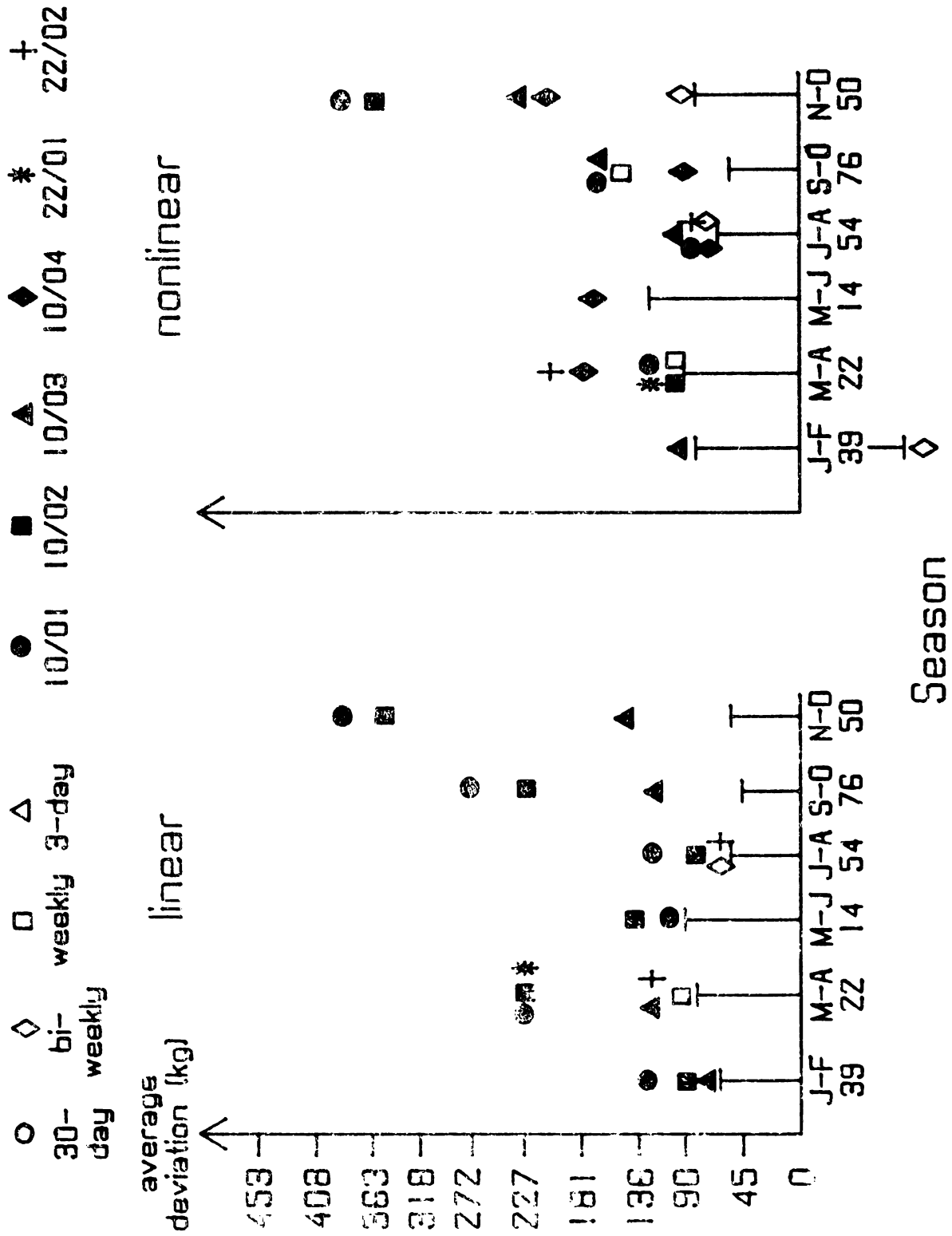
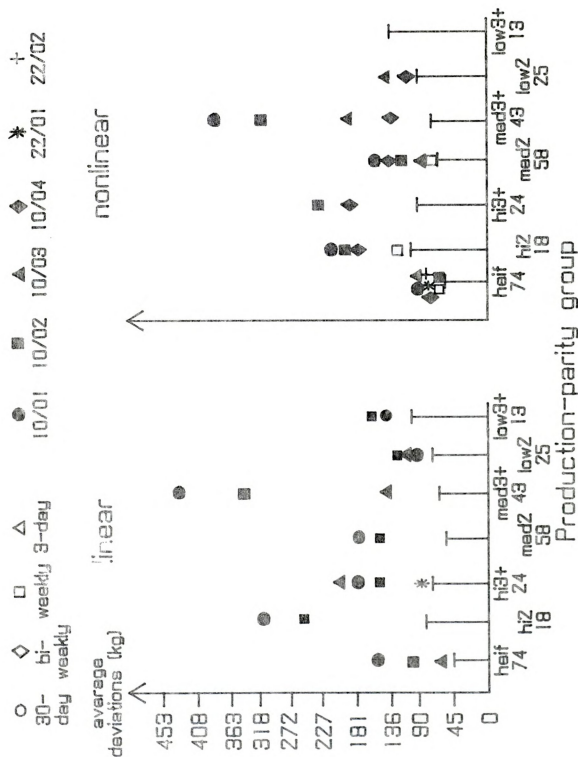
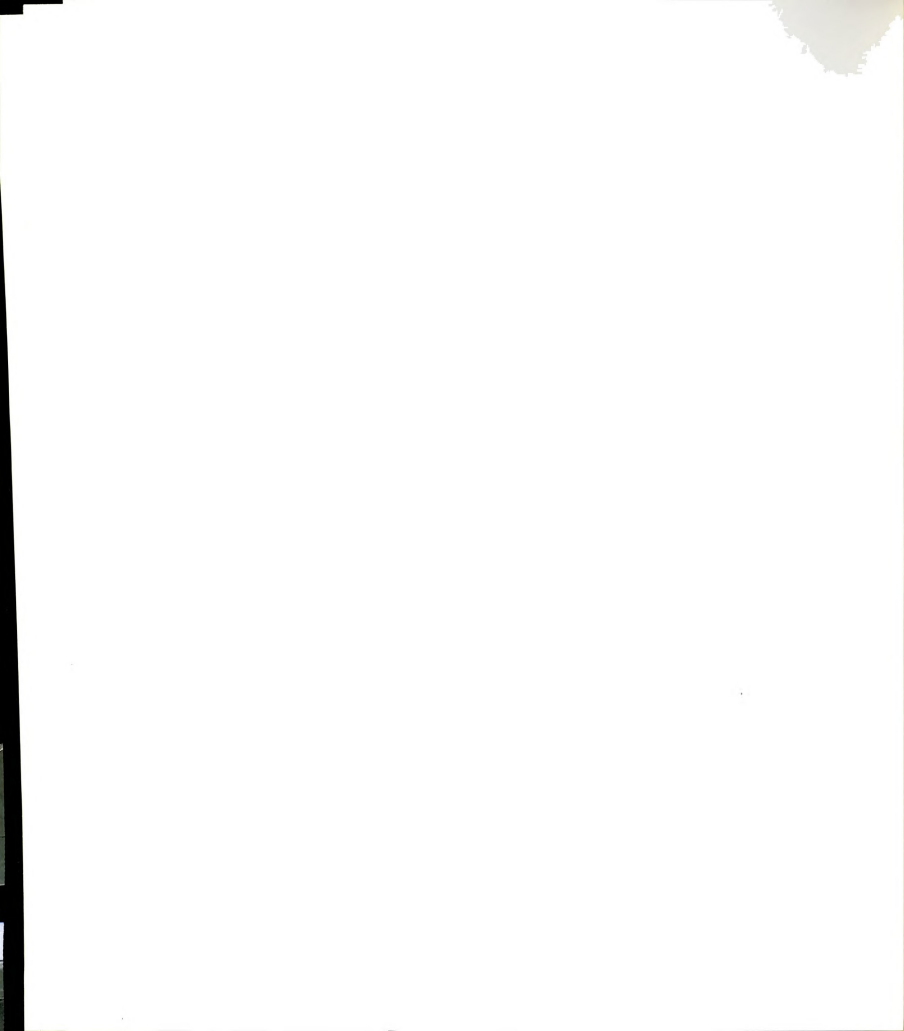




Figure 3. Biases in total yield for production-parity group.







calculated using nonlinear procedure is smaller than for the linear procedure, but the advantage exceeds 50 kg. only for methods with 22 samples.

Table 4. Comparison of root mean squares for ten sampling methods.

| Method   | linear<br>estimation |      | nonlinear<br>estimation |      |
|----------|----------------------|------|-------------------------|------|
|          | kgs.                 | rank | kgs.                    | rank |
| EQUAL    |                      |      |                         |      |
| 30-day   | 171                  | 3    | 164                     | 6    |
| biweekly | 181                  | 4    | 108                     | 3    |
| weekly   | 80                   | 2    | 75                      | 2    |
| 3-day    | 45                   | 1    | 38                      | 1    |
| UNEQUAL  |                      |      |                         |      |
| 10/01    | 332                  | 10   | 288                     | 10   |
| 10/02    | 331                  | 9    | 284                     | 9    |
| 10/03    | 214                  | 7    | 198                     | 8    |
| 10/04    | 211                  | 6    | 167                     | 7    |
| 22/01    | 235                  | 8    | 152                     | 5    |
| 22/02    | 207                  | 5    | 117                     | 4    |

#### 3.3.4 Yield regressed on parameters

The effect of the accuracy of parameter estimates in Wood's equation [1] on the ten sampling methods was examined





by regressing deviations of estimated yield from actual on deviations of estimations from parameters. Results are shown in Table 5.

Table 5. Results of regressing deviations of estimated yield from actual on deviations of estimators from parameters.

|          | standardized partial<br>regression coefficient |                     |                     | observed sig. level<br>partial F stats. |                     |                     |
|----------|------------------------------------------------|---------------------|---------------------|-----------------------------------------|---------------------|---------------------|
| Method   | $\hat{a}-\tilde{a}$                            | $\hat{b}-\tilde{b}$ | $\hat{c}-\tilde{c}$ | $\hat{a}-\tilde{a}$                     | $\hat{b}-\tilde{b}$ | $\hat{c}-\tilde{c}$ |
| EQUAL    |                                                |                     |                     |                                         |                     |                     |
| 30-day   | .92                                            | 1.20                | -.61                | ***                                     | ***                 | ***                 |
| biweekly | .60                                            | .61                 | -.24                | .0002                                   | .0074               | .0764               |
| weekly   | .63                                            | .70                 | -.33                | ***                                     | ***                 | ***                 |
| 3-day    | .55                                            | .86                 | -.85                | ***                                     | ***                 | ***                 |
| UNEQUAL  |                                                |                     |                     |                                         |                     |                     |
| 10/01    | .32                                            | 1.69                | -2.06               | ***                                     | ***                 | ***                 |
| 10/02    | .28                                            | 1.67                | -2.12               | ***                                     | ***                 | ***                 |
| 10/03    | .55                                            | 1.88                | -1.90               | ***                                     | ***                 | ***                 |
| 10/04    | .38                                            | 1.54                | -1.77               | ***                                     | ***                 | ***                 |
| 22/01    | .74                                            | 2.16                | -1.44               | ***                                     | ***                 | ***                 |
| 22/02    | .73                                            | 1.41                | -.66                | ***                                     | ***                 | ***                 |

\*\*\* P < .0001



All of the partial F statistics were significant at  $P < 0.01$ , except one at  $P < 0.08$ . Comparison of the standardized partial regression coefficients (beta weights) for all of the methods with equally-spaced intervals or with at least 22 sample points, indicates that the  $\underline{b}$  parameter, related to the peak of production, ranks first in impact on yield deviations. In all cases it accounts for about 40 to 50% of the importance. For methods with unequal intervals and only ten samples, the  $\underline{c}$  parameter is the most important, indicating that values beyond peak lactation carry about half of the importance. This result may be due to the fact that there is very little sampling after peak production in the associated methods, so that information obtained there becomes more critical in determining the non-linear fit.

#### 3.3.5 Comparisons from MANOVA

Results of multivariate analysis of variance on deviations of estimates from parameters in Wood's equation, using the same fixed factors as previously, and the results of three univariate analyses on the same differences are shown in Table 6. Sampling method interactions with season and production-parity group are significant in all four analyses. Some factors, such as season, are significant in the multivariate results, but not significant in all of the univariate results. One can associate these factor's



effects with stage of lactation by using the appropriate response variable.

Table 6. MANOVA comparisons for deviations from parameters.

| Model factor                | df   | observed significance level |                     |                     |        |
|-----------------------------|------|-----------------------------|---------------------|---------------------|--------|
|                             |      | $\hat{a}-\tilde{a}$         | $\hat{b}-\tilde{b}$ | $\hat{c}-\tilde{c}$ | multi. |
| season                      | 5    | .1346                       | .1439               | .0047               | ***    |
| herd(season)                | 25   | .0263                       | .0958               | .0274               | .012   |
| treatment                   | 3    | .9893                       | .9786               | .9897               | .621   |
| production-<br>parity group | 6    | .0004                       | .1303               | .0555               | ***    |
| days open                   | 1    | .0717                       | .0378               | .0027               | .1411  |
| days in milk                | 1    | .5356                       | .6827               | .4691               | ***    |
| error 1 (cow)               | 213  |                             |                     |                     |        |
| method                      | 9    | .0412                       | .0091               | .0037               | ***    |
| method*season               | 45   | .0512                       | .0002               | .0001               | ***    |
| method*p-p                  | 54   | .0013                       | .0158               | .0013               | ***    |
| error 2 (res)               | 2187 |                             |                     |                     |        |

\*\*\* P < .0001

For example, evidence for the effect of season is marginal when one considers the parameter associated with peak production (b), whereas it is highly significant for the tail parameter (c). The covariate, days in milk, is not seen as an important factor until its effect on all



parameters is considered simultaneously in the MANOVA results.

### 3.3.6 Biases in parameter estimates

Methods that produced parameter deviations not significantly different from zero for each season and for each production-parity group are listed in Tables 7 and 8, respectively. For the initial parameter a in Wood's equation [1], all sampling methods produced biased estimates for all but two seasons and all but one production-parity group, all cases involving sparse data. For parameter b, the sampling methods produced biased estimates for season in 77% of the cases and in 84% of the cases for production-parity groups. Estimation of the c parameter is least affected by sampling, significant biases occurring in fewer than half of the cases. Summarizing the results for all three parameters, methods based on ten sample points often create more bias than do methods based on more sample points, whether those are equally or unequally spaced.



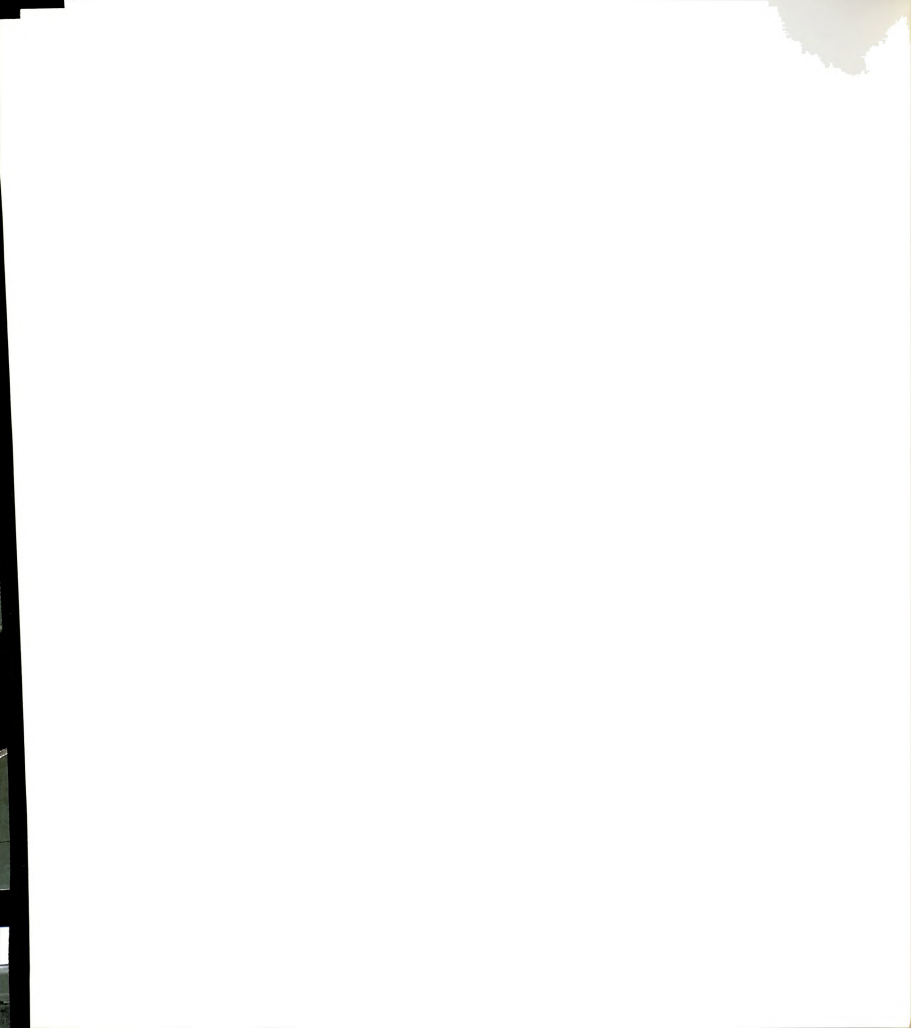


Table 7. Methods that produced parameter deviations which are NOT significantly different from zero, by season.

| parameter<br>estimated | Season                                  |               |                         |                                  |                        |              |
|------------------------|-----------------------------------------|---------------|-------------------------|----------------------------------|------------------------|--------------|
|                        | J-F<br>(39)                             | M-A<br>(22)   | M-J<br>(14)             | J-A<br>(54)                      | S-O<br>(76)            | N-D<br>(50)  |
| <u>a</u>               |                                         | 3-da          | bi-w                    |                                  |                        |              |
| <u>b</u>               | 30-da<br>bi-w<br>week<br>10/01<br>22/02 | 3-da<br>22/02 | bi-w<br>week<br>3-da    | 10/02                            | 22/01                  | bi-w<br>week |
| <u>c</u>               | all                                     | 3-da<br>22/02 | equal<br>22/01<br>22/02 | 30-da<br>10/01<br>10/02<br>22/02 | bi-w<br>22/01<br>22/02 | bi-w<br>week |

equal: all equally spaced sampling methods

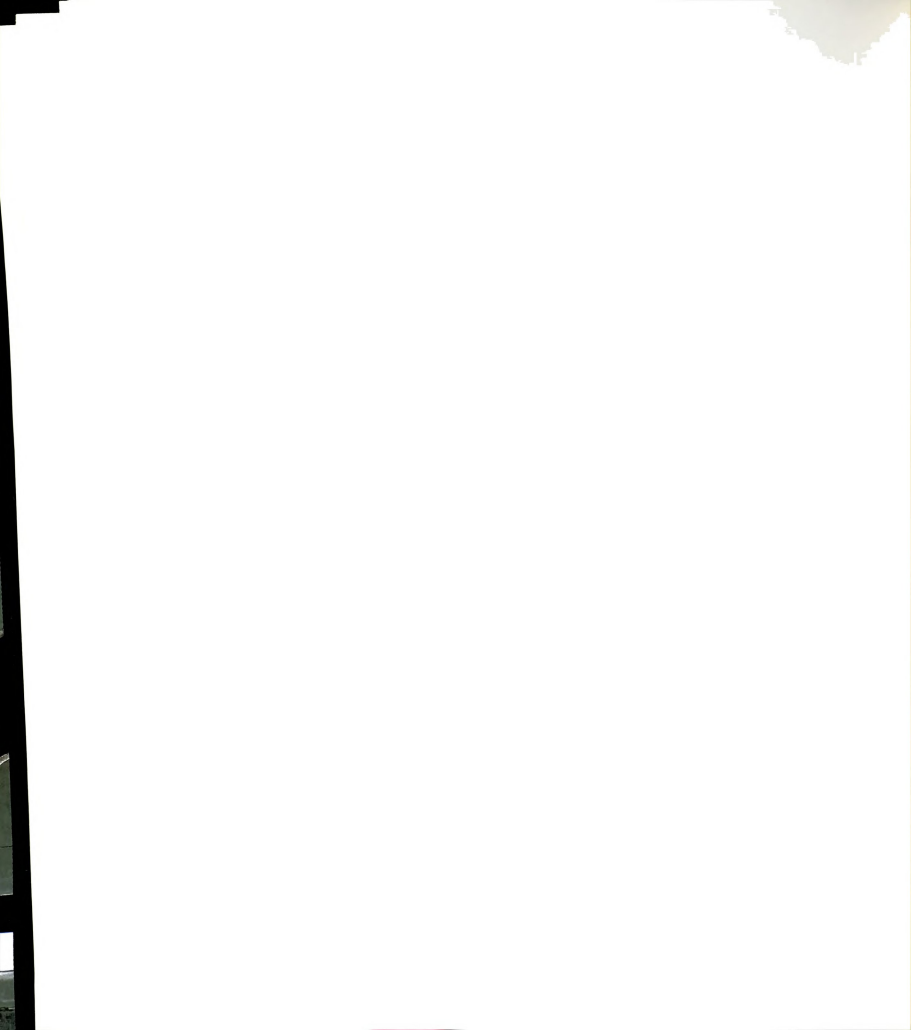
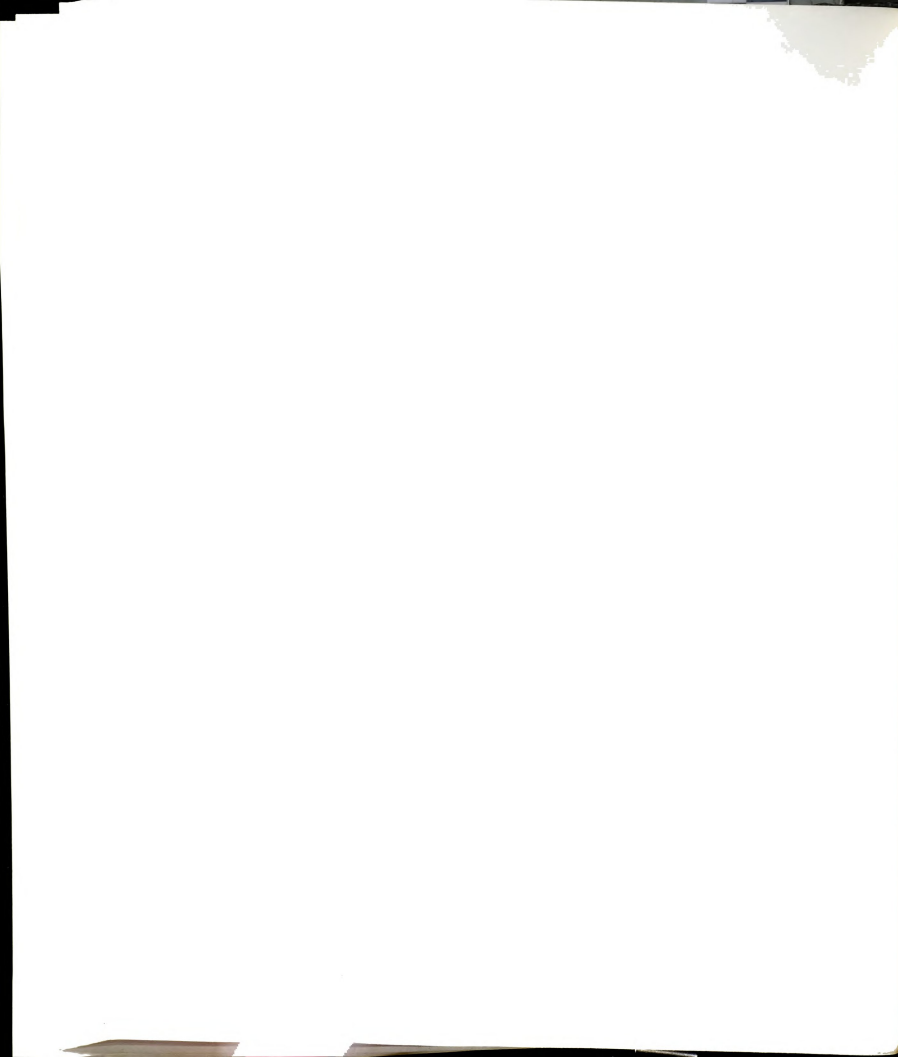


Table 8. Methods that produced parameter deviations NOT significantly different from zero, by production-parity groups.

| Production-parity Group |                         |                      |                                        |              |                        |                                        |                                                    |
|-------------------------|-------------------------|----------------------|----------------------------------------|--------------|------------------------|----------------------------------------|----------------------------------------------------|
| par.<br>est.            | heif<br>(74)            | hi2<br>(18)          | hi3<br>(24)                            | med2<br>(58) | med3<br>(43)           | low2<br>(25)                           | low3<br>(13)                                       |
| <u>a</u>                |                         |                      |                                        |              |                        |                                        | equal<br>10/02<br>10/04<br>22/01<br>22/02          |
| <u>b</u>                |                         |                      | bi-w                                   |              |                        |                                        | equal<br>10/01<br>10/02<br>10/04<br>22/01<br>22/02 |
| <u>c</u>                | 10/02<br>22/01<br>22/02 | bi-w<br>week<br>3-da | bi-w<br>week<br>3-da<br>22/01<br>22/02 |              | 3-da<br>22/01<br>22/02 | bi-w<br>week<br>3-da<br>22/01<br>22/02 | equal<br>10/01<br>10/04<br>22/01<br>22/02          |

equal: all equally spaced sampling methods



### 3.3.7 Precision of parameter estimates

Root mean square residual errors for estimation of nonlinear parameters are listed by sampling method in Table 9. The parenthetic numbers are the ranks of sampling methods for a given parameter. Across parameters, there is almost complete agreement of ranks by method, i.e., the relative precision of one method is similar across the three major stages of lactation.

**Table 9.** Comparison of root mean squares of parameter deviations for ten sampling methods.

| Method   | $\hat{a}-\tilde{a}$ |      | $\hat{b}-\tilde{b}$ ( $\times 10^{-2}$ ) |      | $\hat{c}-\tilde{c}$ ( $\times 10^{-4}$ ) |      |
|----------|---------------------|------|------------------------------------------|------|------------------------------------------|------|
|          | kgs.                | rank | kgs.                                     | rank | kgs.                                     | rank |
| EQUAL    |                     |      |                                          |      |                                          |      |
| 30-day   | 7.69                | (6)  | 5.34                                     | (5)  | 6.93                                     | (5)  |
| biweekly | 5.84                | (4)  | 3.54                                     | (3)  | 3.66                                     | (3)  |
| weekly   | 5.47                | (2)  | 3.25                                     | (1)  | 3.34                                     | (2)  |
| 3-day    | 5.74                | (3)  | 3.32                                     | (2)  | 3.14                                     | (1)  |
| UNEQUAL  |                     |      |                                          |      |                                          |      |
| 10/01    | 9.35                | (8)  | 7.95                                     | (9)  | 12.97                                    | (9)  |
| 10/02    | 9.77                | (10) | 8.01                                     | (10) | 13.42                                    | (10) |
| 10/03    | 8.58                | (7)  | 6.69                                     | (7)  | 9.61                                     | (7)  |
| 10/04    | 9.57                | (9)  | 7.46                                     | (8)  | 10.55                                    | (8)  |
| 22/01    | 6.92                | (5)  | 5.51                                     | (6)  | 7.39                                     | (6)  |
| 22/02    | 5.18                | (1)  | 4.26                                     | (4)  | 5.29                                     | (4)  |



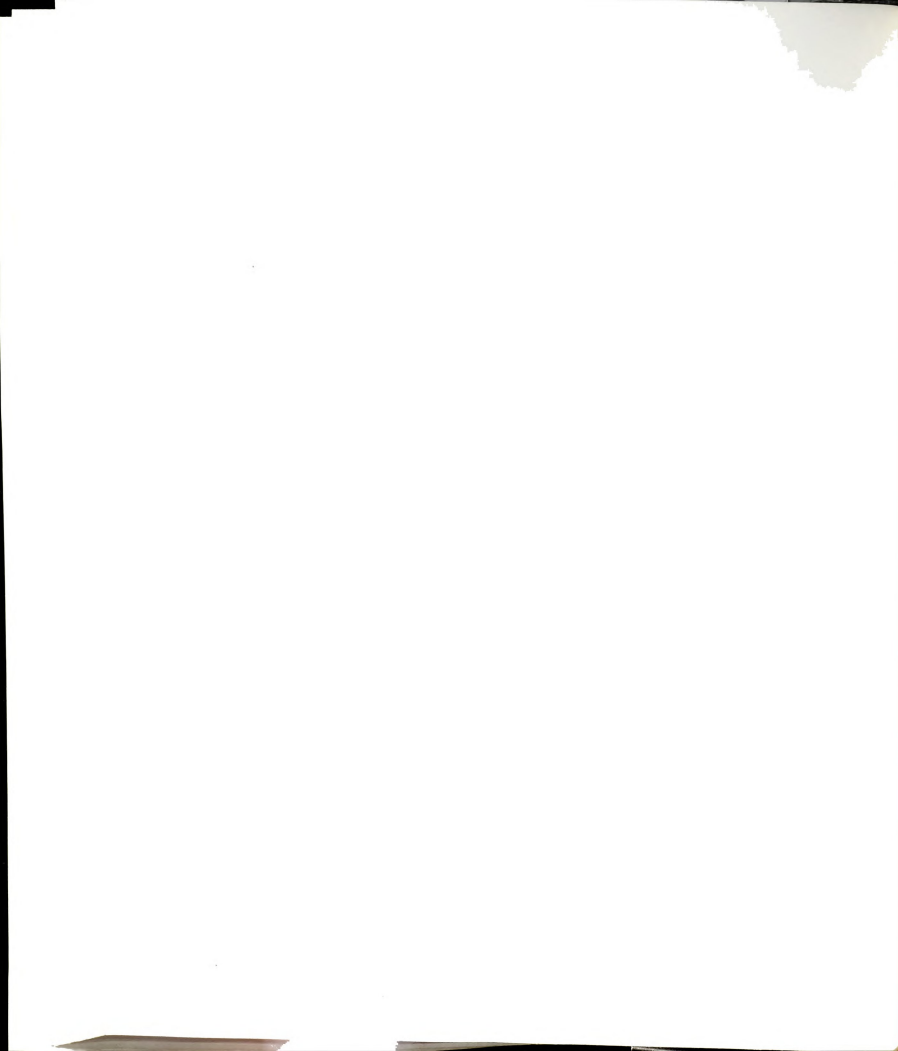
### 3.4 Conclusions

Sampling methods with four or fewer observations after the peak of lactation exhibit more bias and are less precise than methods that include more than four. Almost all of the sampling methods over-estimated actual yield to varying degrees. Bias in estimation error increases as the length of the sampling interval increases after the peak of lactation, and this bias cannot be ameliorated by decreasing the length of the sampling interval before and during the peak of lactation. In regard to the model itself, the interactions of sampling method with season and with production-parity group are significant factors as is herd within season. Results obtained from a model with a covariate, days in milk, are likely to be different than results obtained with the covariate removed from the model.

If a researcher intends to use only the test interval method (our linear procedure) to estimate total yield, there is little need to sample more often than every 30 days. One unequally spaced method (ten samples, with four after peak) produces estimates of yield not significantly different from the 30-day method (ten samples, with seven after peak) and has the advantage of ending the sampling 46 days earlier.

If, on the other hand, an investigator wants to use Wood's equation (our nonlinear procedure) to estimate milk yield, then all unequally-spaced methods based on ten samples are significantly biased, whereas the 30-day





equally-spaced scheme is not. Estimates based on 22 unequally-spaced observations are biased in some seasons and/or production-parity groups. For methods that concentrate sampling before and during peak lactation, Wood's  $c$  parameter (representing post-peak decline) exhibits the most influence on estimated yield. For methods that sample equally throughout lactation,  $b$  (the slope to peak parameter) is most important.

None of the sampling methods, not even the one with 3-day intervals, provides an adequate estimation procedure for the parameters in Wood's equation. A comparison of multivariate versus univariate analyses on deviations of estimates from parameters indicates that some fixed factors are more important during various stages of lactation. But, to obtain an overall picture of lactation, one must consider all of the factors and the covariates simultaneously.



#### 4. SUMMARY

The first objective of this investigation was to determine the effect of unequally spaced sampling intervals on the accuracy and precision of estimating total lactation yield. There were six unequally spaced sampling methods. They gradually increased observations taken during the time of maximum (peak) production which is also the time of greatest variation. To permit comparison with standard practice, daily records were also sampled at equally spaced intervals of different lengths.

It is known that factors other than when and how often a cow is sampled affect the estimate of total yield. The analysis of a linear model with both fixed and continuous factors indicated that sampling produces different estimates of yield for various seasons and for different production-parity groups. It was also noted that when the non-significant covariate, days in milk, was removed from the model other factors became significant.

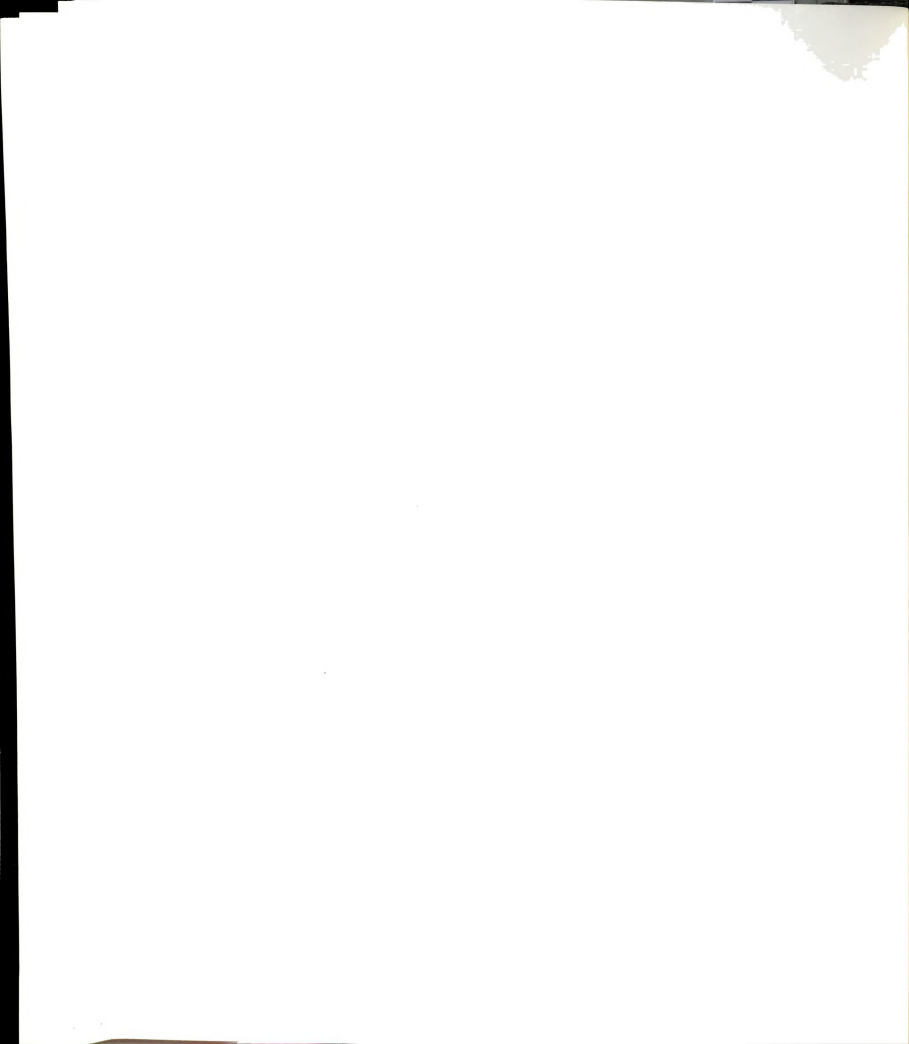
The results of the analysis of variance also indicate that sampling methods with four or fewer observations after the peak of production exhibit more bias and are less



precise than methods that include more than four. Almost all of the sampling methods overestimated actual yield. When yield was underestimated, it was within the margin of error in all but one instance. Finally, bias in estimation increases as the length of the sampling interval increases after the peak of lactation. This bias cannot be ameliorated by decreasing the length of the sampling interval before and during peak production.

Each sampling method was used to obtain two estimates of actual yield to determine what effect method had on the estimation procedure itself. The first estimation procedure (linear estimation) revealed that sampling once every 30 days produces acceptable estimates as compared to samples taken at least twice as often. The second (nonlinear) procedure for estimation produced biased estimates of yield for all unequally spaced intervals based on ten observations. Estimates based on 22 unequally spaced observations are biased in some seasons and some production-parity groups.

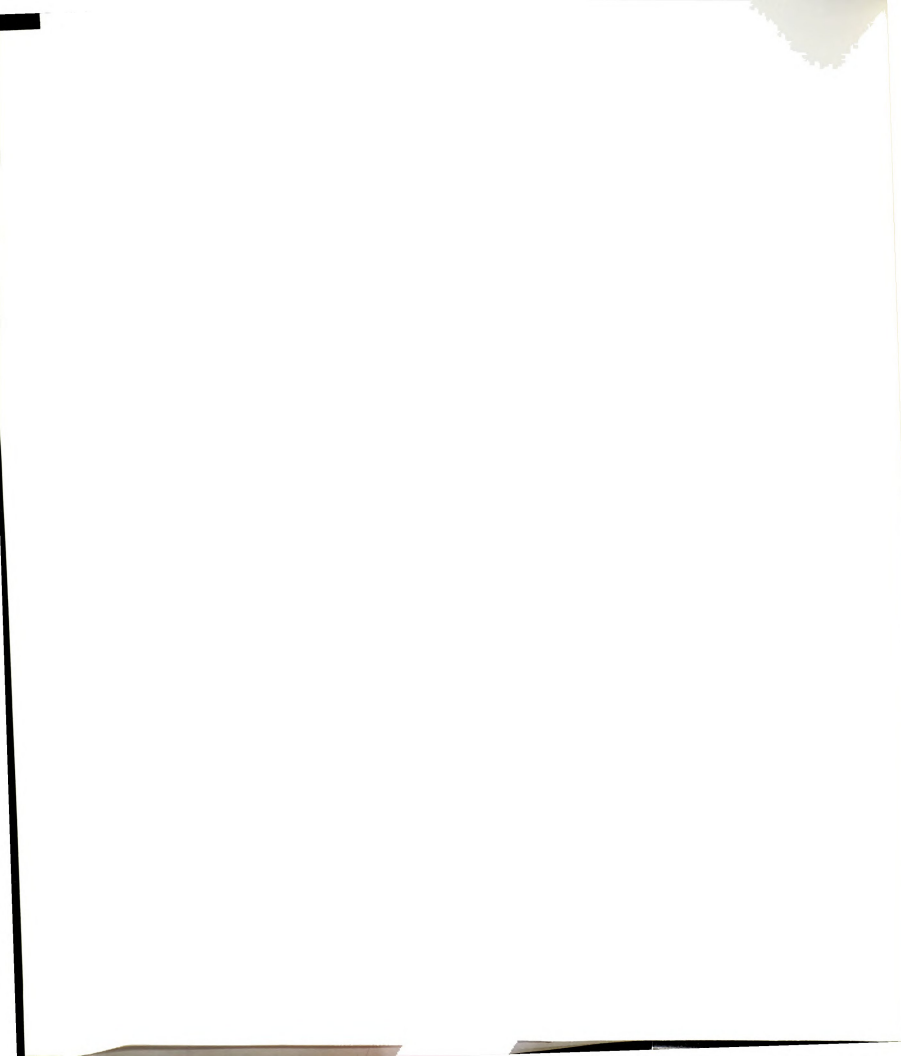
The second objective was to determine the accuracy of nonlinear estimation of various characteristics of lactation curves on the accuracy of estimating lactation yield. The nonlinear equation proposed by Wood (1967) was used for all of the sampling methods and estimates of the parameters were obtained. For methods that concentrate sampling before and during peak production, Wood's  $c$  parameter (representing



post-peak decline) exhibits the most influence on estimated yield. For methods that sample equally throughout lactation,  $b$  (the slope to peak parameter) is most important.

The last objective was to determine the effect of the factors in the linear model on the accuracy and precision of the estimates of parameters in Wood's equation [1]. Some factors are significant in multivariate analysis when all three parameters are considered simultaneously, but not significant in one or more of the univariate results. One can associate the effect of these factors with the characteristic of the lactation curve by using the appropriate response variable. Finally, these results indicate that parameter estimates are usually biased regardless of the sampling method, but precision of their estimates increases as the length of the interval decreases.





## BIBLIOGRAPHY

- Alexander, M.H., and W.W. Yapp. 1949. Comparison of methods of estimating milk and fat production in dairy cows. *J. Dairy Science* 32:621.
- Anderson, C.R. 1981. A biometrical and genetic study of Tribolium egg production curves as a model for lactation curves. Ph.D. thesis University of Illinois Urbana-Champaign, Illinois.
- Badner, G.B., C.R. Anderson, I.L. Mao, and J.P. Walter. 1984. A comparison of lactation curves fitted from daily, weekly, biweekly, and monthly milk weights. *J. Dairy Science* 67 Abstracts:182.
- Badner, G.B., and C.R. Anderson. 1985. Evaluation of five lactation curve models fitted from daily milk weights. *J. Dairy Science* 68 Abstracts:226.
- Barta, T.R., and A.J. Lee. 1985. Comparison of three methods of predicting 305-day milk and fat production in dairy cows. *Canadian J. of Animal Science* 65:341.
- Cobby, J.M., and Y.L.P. LeDu. 1978. On fitting curves to lactation data. *Animal Production* 26:127.
- Cole, J.W.L., and J.E. Grizzle. 1966. Applications of multivariate analysis of variance to repeated measurements experiments. *Biometrics* 22:810.
- Congleton, W.R. Jr., and R.W. Everett. 1980a. Error and bias in the incomplete gamma function to describe lactation curves. *J. Dairy Science* 63:101.
- Congleton, W.R. Jr., and Everett, R.W. 1980b. Application of the incomplete gamma function to predict cumulative milk production. *J. Dairy Science* 63:109.
- Cunningham, E.P., and V.E. Vial. 1968. Relative accuracy of different sampling intervals and methods of estimation for lactation milk yield. *Irish J. Agricultural Research* 7:49.

Everett, R.W., H.W. Carter, and J.D. Burke. 1968. Evaluation of the Dairy Herd Improvement Association record system. *J. Dairy Science* 51:153.

Everett, R.W., B.T. McDaniel, and H.W. Carter. 1968. Accuracy of monthly, bimonthly, and trimonthly Dairy Herd Improvement Association Records. *J. Dairy Science* 51:1051.

Everett, R.W., and H.W. Carter. 1968. Accuracy of test interval method of calculating Dairy Herd Improvement Association records. *J. Dairy Science* 51:1936.

Ferris, T.A. 1981. Selecting for lactation curve shape and milk yield in dairy cattle. Ph.D. thesis. Michigan State University East Lansing, Michigan.

Geisser, S. 1963. Multivariate analysis of variance for a special covariance case. *J. American Statistical Association* 58:660.

Gill, J.L. 1978. *Design and Analysis of Experiments in the Animal and Medical Sciences Volume 1* The Iowa State University Press. Ames, Iowa.

Greenhouse, S.W., and Geisser, S. 1959. On methods in the analysis of profile data. *Psychometrika* 24:95.

Kellogg, D.W., N.S. Urquhart, and A.J. Ortega. 1977. Estimating Holstein lactation curves with a gamma curve. *J. Dairy Science* 60:1308.

Keown, J.K., R.W. Everett, N.B. Emptet, and L.H. Wadell. 1986. Lactation curves. *J. Dairy Science* 69:769.

Marquardt, D.W. 1963. An algorithm for least-squares estimation of nonlinear parameters. *J. of the Society of Industrial Applied Mathematics* 1:431.

McDaniel, B.T. 1969. Accuracy of sampling procedures for estimating lactation yields: a review. *J. Dairy Science* 52:1742.

McNally, D.H. 1971. Mathematical models for poultry egg production. *Biometrics* 27:735.

Menchaca, M.A. 1981. Comparison of estimation methods and sampling intervals in the milk yield prediction. *Cuban J. Agricultural Science* 15:1.

- Miller, P.D., W.E. Lentz, and C.R. Henderson. 1970. Joint influence of month and age of calving on milk yield of Holstein cows in the Northeastern United States. *J. Dairy Science* 53:351.
- Nelder, J.A. 1966. Inverse polynomials, a useful group of multi-factor response functions. *Biometrics* 22:128.
- O'Connor, L.K. and S. Lipton. 1960. The effect of various sampling intervals on the estimation of lactation milk yield and composition. *J. Dairy Research* 27:389.
- Sargent, F.D., V.H. Lytton, and O.G. Wall, Jr. 1968. Test interval method of calculating Dairy Herd Improvement records. *J. Dairy Science* 51:170.
- S.A.S. Institute Inc. 1985. *S.A.S. User's Guide: Statistics Version 5 Edition*. S.A.S Institute Inc. Cary, North Carolina.
- Schaeffer, L.R., and E.B. Burnside. 1976. Estimating the shape of the lactation curve. *Canadian J. Animal Science* 56:157.
- Schaeffer, L.R., C.E. Minder, I. McMillan, and E.B. Burnside. 1977. Nonlinear techniques for predicting 305-day lactation production of Holstein and Jerseys. *J. Dairy Science* 60:1636.
- Shook, G.E., L.P. Johnson, and F.N. Dickinson. 1980. Factors for improving accuracy of estimates of test-interval yield. *Dairy Herd Improvement Letter, Volume 56, Number 4*. United States Department of Agriculture Science and Education Administration Beltsville, Maryland.
- Switzky, D. 1985. Shooting for a high peak yield. *Dairy* September, 1985.
- Wiggans, G.R., and M. Grossman. 1980. Computing lactation records from sample-day production. *Dairy Herd Improvement Letter, Volume 56, Number 4*. United States Department of Agriculture Science and Education Administration Beltsville, Maryland.
- Wood, P.D.P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164.
- Wood, P.D.P. 1969. Factors affecting the shape of the lactation curve in cattle. *Animal Production* 11:307.

Wood, P.D.P. 1970. A note on the repeatability of parameters of the lactation curve in cattle. *Animal Production* 12:535.

Wood, P.D.P. 1972. A note on seasonal fluctuations in milk production. *Animal Production* 15:89.

Wood, P.D.P. 1976. Algebraic models of the lactation curves for milk, fat, and protein production with estimates of seasonal variation. *Animal Production* 22:35.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03082 2013