# A PROJECT FOR SOME MONTE CARLO STUDIES OF VARIANCE COMPONENT ESTIMATES*

BU-161-M S. R. Searle October, 1963

## Abstract

This paper represents some preliminary notes about Monte Carlo studies that are being contemplated for an investigation into the distribution of variance component estimates obtained from unbalanced data. The presentation is in terms of the 2-way classification, and covers such topics as calculating sampling variances, selecting patterns of $n_{ij}$-values, sampling cell means, sampling the normal distribution and the uniform distribution, and estimating the histogram of the probability distribution of a variance component estimate.

*BU-161-M of the Biometrics Unit Mimeo Series, Cornell University

# A PROJECT FOR SOME MONTE CARLO STUDIES OF VARIANCE COMPONENT ESTIMATES*

BU-161-M S. R. Searle October, 1963

Henderson (1953) gives methods for estimating variance components from unbalanced data. Searle (1956, 1958 and 1961) gives procedures based on the normality assumption for obtaining variances of such estimates in the one-way, two-way and two-way nested classifications, using Method 1 of Henderson's paper. Mahamunulu (1963) uses the same procedures for the three-way nested classification. Monte Carlo studies are now being contemplated to investigate the distribution of variance components estimates obtained by Henderson's method, utilizing the variances of the estimates derived by Searle's procedures. These notes are a preliminary discussion of some of the methods that are being considered. They are presented in terms of the two-way (cross) classification.

## 1. Calculating Sampling Variances

### The n-pattern

Suppose in a two-way classification of a rows and b columns there are $n_{ij}$ observations in the cell defined by the $i^{th}$ row and $j^{th}$ column. Let $n_{i.}$ be the total number of observations in the $i^{th}$ row and $n_{.j}$ the total in the $j^{th}$ column, with $n_{..}$ altogether. Thus

No. of observations

| Row | Column | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | ... | j | .... | b | |
| 1 | | | | | | | |
| 2 | | | | | | | |
| ⋮ | | | | | | | |
| i | | | | $n_{ij}$ | | | $n_{i.}$ |
| ⋮ | | | | | | | |
| a | | | | | | | |
| Total | | | | $n_{.j}$ | | | $n_{..}$ |

*BU-161-M of the Biometrics Unit Mimeo Series, Cornell University

We will refer to any set of values of the $n_{ij}$'s, $n_{i.}$'s $n_{.j}$'s as an n-pattern.

Estimates

If $T_a$ = uncorrected sum of squares for rows
$T_b$ = " " " " " columns
$T_{ab}$ = " " " " " interaction
$T_o$ = " total sum of squares
$T_f$ = correction factor

then, as suggested in Searle (1958), the estimated variance components can be expressed as

$$\left.\begin{aligned} T_a - T_f &= q_1\hat{\sigma}_r^2 + q_2\hat{\sigma}_c^2 + q_3\hat{\sigma}_{rc} + q_4\hat{\sigma}_e^2 \\ T_b - T_f &= q_5\hat{\sigma}_r^2 + q_6\hat{\sigma}_c^2 + q_7\hat{\sigma}_{rc} + q_8\hat{\sigma}_e^2 \\ T_{ab} - T_a - T_b + T_f &= q_9\hat{\sigma}_r^2 + q_{10}\hat{\sigma}_c^2 + q_{11}\hat{\sigma}_{rc} + q_{12}\hat{\sigma}_e^2 \end{aligned}\right\} \quad (1)$$

$$T_o - T_{ab} = q_{13}\hat{\sigma}_e^2 \ .$$

The q's are constants, functions of the n-pattern. They are such that an explicit solution of these equations for the $\hat{\sigma}^2$'s is not feasible algebraically. But for a given set of n's the q's can be computed and the equations solved as

$$\left.\begin{aligned} \hat{\sigma}_r^2 &= \lambda_1 T_a + \lambda_2 T_b + \lambda_3 T_{ab} + \lambda_4 T_f + p_1\hat{\sigma}_e^2 \\ \hat{\sigma}_c^2 &= \lambda_5 T_a + \lambda_6 T_b + \lambda_7 T_{ab} + \lambda_8 T_f + p_2\hat{\sigma}_e^2 \\ \hat{\sigma}_{rc}^{\ 2} &= \lambda_9 T_a + \lambda_{10} T_b + \lambda_{11} T_{ab} + \lambda_{12} T_f + p_3\hat{\sigma}_e^2 \ , \end{aligned}\right\} \quad (2)$$

where the $\lambda$'s and p's are appropriate constants.

Variances of estimated components

For any given n-pattern the q's of the first set of equations and the $\lambda$'s and p's of the second set can be calculated. The estimate of the within sub-class variance, $\hat{\sigma}_e^2$, is independent of the T's; and Searle (1958) gives expressions for the variances and covariances of the T's. Variances of variance component estimates derived from equation (2) can therefore be obtained. Algebraic complexity prevents explicit expressions being obtained for these variances, which involve both the n-pattern and the true variance components. But for a given n-pattern and a given set of values of the true variance components the variances can be calculated. These will be the true (sampling) variances of variance component estimates derived from data with the given n-pattern that are a sample from a population having the given set of values as true variance components.

Prior to Monte Carlo sampling of such a situation we will first make a series of calculations to investigate properties of these sampling variances. For a given n-pattern, the sampling variances will be computed for various sets of values of the true variance components - and for each set of such components the sampling variances will be computed for various n-patterns. This study will be of interest in itself, and among other things it may, hopefully, indicate which n-patterns and which sets of values for true variance components would be worthy of use in the Monte Carlo sampling. If the sampling variances are relatively insensitive to small changes in the n-patterns or in the values used for the true variance components, then only combinations of these that lead to sampling variances substantially different in value will be used for the Monte Carlo sampling. But the first problem is to select the n-patterns.

### Selecting n-patterns

Initially one might contemplate starting this project with only a small number of rows and columns and using n-patterns that are patently different and easily specified: e.g. in a configuration of 4 rows and 5 columns two such n-patterns might be

$$\begin{array}{ccccc} n & n & 0 & 0 & 0 \\ 0 & n & n & 0 & 0 \\ 0 & 0 & n & n & 0 \\ 0 & 0 & 0 & n & n \end{array} \qquad \text{and} \qquad \begin{array}{ccccc} n & n & 0 & 0 & 0 \\ n & n & 0 & 0 & 0 \\ n & n & n & n & n \\ n & n & n & n & n \end{array} \quad .$$

Patterns such as these have been studied by Bush and Anderson (1962); they are certainly different, one from another, and they are easily specified and simple to describe - but whereas they may be of some interest in analysing data from a fixed-effects model they bear little relation to some of the practical problems that arise in estimating variance components from random-effects models. One particular instance where this estimation procedure is widely used is that of dairy production records of cows in various herds sired by a (relatively) few bulls used in artificial breeding. In this case columns may represent bulls and rows represent herds, and an analysis of actual data may well contain as many as 100 columns and 600 rows. Furthermore, maybe only 10% of all cells will have observations in them and of these maybe 70% have only one observation. This situation is so far removed from the well-organized n-patterns illustrated above that it is felt that a study of such patterns would give little information about the practical problem. Possible methods of considering the latter have therefore been evolved.

Considering for a moment the dairy production problem we may note that the distribution of a sire's daughter's over the population of herds using A.B. is operationally fairly much a random process. Furthermore, empirical distributions of the number of daughters from a given sire in a given herd can be easily obtained. One such is given in Table 1. This in, effect, is the distribution of $n_{ij}$-values in a study of actual data made recently in New Zealand (Searle, 1963). One way of establishing an n-pattern for an analysis of a rows and b columns would be to take for each row (representing a herd) b randomly chosen values from the empirical distribution given in Table 1 (or one like it, based on other data - from New York for example).

There is, however, a disadvantage to the above suggestion. The number of observations in a row total, $n_{i.}$, (i.e. herd size) would then depend largely on the number of columns (sires) in the analysis. And this is not true in practice - an A.B. stud has enough bulls at any one time to cater to the whole of its clientele, but the size of an individual farmer's herd is certainly not determined by the number of bulls at the A.B. stud. This mitigates against the method suggested above for setting up n-patterns.

The number of cows that a farm has is largely a matter of specific limitation by the farmer concerned, and this factor must be taken into account. This can be done by deciding on a series of values for the number of observations in a row (herd) and then distributing them randomly among the columns (bulls) according to the empirical distribution in Table 1. The values for the number of observations in a row can themselves be selected from a distribution of herd sizes, such as is given in Table 2. For a rows, a values will be chosen; these will be the $n_{i.}$ values. Then, for each row, sufficient non-zero values could be chosen at random from the distribution given in Table 1 such that their total is $n_{i.}$. These would be the non-zero $n_{ij}$ values for that row, and together with the appropriate number of zeros these would be distributed at random among the b columns of the row. And so for every row. The $n_{.j}$ values would then simply be the sum of the appropriate $n_{ij}$'s.

To better specify the distributions in Tables 1 and 2 it may be possible to fit Poisson distributions to them. Even if the fit is not statistically a very good one, so long as it is not horrendously bad this may be a suitable technique. It would provide an opportunity for easily specifying other (Poisson) distributions.

If Poisson distributions are suitable in the above discussion further refinement is possible. First use a Poisson distribution to establish the numbers of observations in a row, the $n_{i.}$ values. Then, knowing that

$$p(n_{i.}) \text{ is Poisson}$$

and

$$p(n_{ij}) \text{ is Poisson}$$

the distribution

$$p(n_{i1}\ n_{i2}\ \ldots\ n_{ib} \mid n_{i.}) \text{ is Multinomial}$$

and can be sampled for the $n_{ij}$'s for each row. For example, in the above Tables the mean value of $n_{i.}$ is 7546/654 = 11.54.

Suppose

$$m = 11.54$$

and

$$p(n_{i.}) = \frac{e^{-m} m^{n_{i.}}}{\underline{|n_{i.}}} \ .$$

Then, with 119 sires the mean value of $n_{ij}$ (including zeros) is

$$\frac{m}{119} = \frac{11.54}{119} = 0.09696 \ .$$

Thus we will suppose that

$$p(n_{ij}) = \frac{e^{-m/119}(m/119)^{n_{ij}}}{\underline{|n_{ij}}} \ .$$

Then

$$p(n_{i1}\ n_{i2} \cdots n_{i,119} \mid n_{i.}) = \frac{\prod\limits_{j=1}^{119} p(n_{ij})}{p(n_{i.})}$$

$$= \frac{e^{-m}(m/119)^{n_{i.}}}{\prod\limits_{j} \left|\underline{n_{ij}}\right.} \Bigg/ \frac{e^{-m}m^{n_{i.}}}{\left|\underline{n_{i.}}\right.}$$

$$= \frac{\left|\underline{n_{i.}}\right.}{\left|\underline{n_{i1}}\right. \left|\underline{n_{i2}}\right. \cdots \left|\underline{n_{i119}}\right.} \left(\frac{1}{119}\right)^{n_{i.}}$$

which is a multinomial distribution with all p's equal to 1/119, the reciprocal of the number of sires (columns). One must, of course, test the fit of these Poisson distributions to the empirical distributions in Tables 1 and 2. It will also be useful to have other empirical distributions, such as those from New York dairy data for example.

Having chosen a method (or methods) for specifying n-patterns for a configuration of a rows and b columns, further control of the n-pattern specification will lie in the number of rows and columns chosen for study. Four easy possibilities are available: few rows and many columns, many rows and few columns, or many or few of both. Having decided on a series of n-patterns - perhaps as many as 20, or 50 or even 100 - (?) - each will be used in conjunction with various sets of values for the true variance components, to calculate sampling variances of variance component estimates. In practice the true components will be taken as fractions of $\sigma_e^2$, then using 1.00 for $\sigma_e^2$ in the computations and values in the range 0.1 to 2.0 - (?) - for the row, column and interaction components.

## 2. Monte Carlo Procedures

### Sampling cell means

Some of the n-patterns and sets of values of the true variance components used in the calculations described above will be used in carrying out the Monte Carlo procedures. Sampling normal populations to obtain cell means will proceed as follows. Suppose $t_i$, $t_j$, $t_{ij}$ and $\bar{t}_{ij.}$ are four randomly-chosen values from a standardized normal distribution (zero mean and unit variance).

If $\sigma_r^2$, $\sigma_c^2$, $\sigma_{rc}^2$ and $\sigma_e^2$ represent the set of values of the true variance components, the "observed" cell mean obtained from the sampling procedure for the cell corresponding to row i and column j will be

$$\bar{x}_{ij.} = t_i\sigma_r + t_j\sigma_c + t_{ij}\sigma_{rc} + \frac{\bar{t}_{ij.}\sigma_e}{\sqrt{n_{ij}}} \quad . \qquad (3)$$

For each cell for which $n_{ij} \neq 0$ a sample of values $t_i$, $t_j$, $t_{ij}$ and $\bar{t}_{ij.}$ will be drawn and $\bar{x}_{ij.}$ calculated. Individual "observations", $x_{ijk}$, will not be sampled. Estimating variance components from these cell means is discussed below, but first some comments on drawing samples from the standardized normal distribution.

Sampling the normal distribution

Several methods of sampling the normal distribution are discussed by Muller (1959). All of them rely on first sampling the uniform distribution, the simplest procedure then being to invoke the central limit theorem and use the mean of a number of variables chosen randomly from a uniform distribution; i.e. the mean of a sample of random numbers chosen from a given interval. The fastest and most accurate procedure is that due to Box and Muller (1958) of using a pair of such random numbers, $U_1$ and $U_2$ say, to generate a pair of random normal deviates by the transformations

$$X_1 = \sqrt{-2 \log_e U_1}\ (\cos 2\Pi\, U_2) \text{ and } X_2 = \sqrt{-2 \log_e U_1}\ (\sin 2\Pi\, U_2) \ .$$

Both methods take a relatively long time to compute, especially when repeated a vast number of times as will be the case in the studies under consideration. An alternative is therefore proposed, based on a table look-up procedure utilizing a random number.

Stored in the computer will be 1000 values, w(s), such that

$$\frac{1}{\sqrt{2\Pi}} \int_0^{w(s)} e^{-t^2/2} dt = \frac{s}{2000}$$

$$\text{for } s = 0, 1, 2, 3, \ldots, 999.$$

On each occasion that a random normal deviate is required two random numbers will be generated:

s, a random number between 000 and 999,

and z, a random sign, +1 or -1.

The corresponding random normal deviate will then be $z[w(s)]$. This will be a fast procedure computationally, and it is felt that the degree of approximation to true normal sampling is likely to be quite sufficient for the studies being undertaken.

It is easily seen that the values $w(s)$ for $s = 0, 1, \ldots, 999$ divide the positive half of the normal distribution into 1,000 equi-probable areas having probability 1/2000. A table of $w(s)$ has been prepared and is available. One might note that it can be used very easily to generate similar tables; e.g. every fourth value yields a table of equi-probable areas having probability 1/500.

## Sampling the uniform distribution

The power residue method of sampling the uniform distribution is well described by the Applied Research Laboratory (1962). The procedure is as follows. To generate a series of integers, each of d digits, $(d > 3)$, start with any such integer $U_0$, that is not divisible by 2 or 5. Choose another integer X, of the form

$$X = 200t \pm r$$

where t is any integer

and r is any one of the values 3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61, 67, 69, 77, 83, 91,

such that X is close to $10^{d/2}$.

The series $U_1, U_2, \ldots$ is now generated by taking the d right-hand digits of the successive products of $U_0$ with X. Thus

$$\begin{aligned} U_1 &= \text{right-hand d digits of } XU_0 \\ U_2 &= \text{" \quad " \quad " \quad " \quad "} \; XU_1, \\ &\vdots \qquad\qquad\qquad\qquad\quad \vdots \\ \text{and} \quad U_n &= \text{" \quad " \quad " \quad " \quad "} \; XU_{n-1}. \end{aligned}$$

This yields a series of numbers that are uniform and random as judged by various statistical tests, see IBM (1959). The series is only pseudo-random in that it repeats itself after $5(10^{d-3})$ terms.

The above method of generating random numbers of 3 digits will result in a series that repeats after 50 values, for any given $U_0$ and X. This is impractical for the purposes needed. But if 7-digit numbers are generated, (the maximum size on a 48-bit word computer) they can be classified by their first three digits for the purposes required; i.e. for all numbers in the range 0000000 to 0009999, s will be taken as 000; for those in the range 0010000 to 0019999, s will be taken as 001; and so on. This will give a series that repeats itself only after 500,000 terms. For each of the random normal deviates required for $\bar{x}_{ije}$, the t's in equation (3), a different pair of values for X and $U_0$ will be used. And if the number of filled cells in an n-pattern is N, new pairs of values for X and $U_0$ will be required after 500,000/N samples have been taken. If fewer samples than this are taken for any particular combination of n-pattern and set of true variances, the values of X and $U_0$ will be different for each such combinations; if more than 500,000/N samples are taken X and $U_0$ will be changed after 500,000/N samples and after each combination of n-pattern and set of true variances.

The distribution of estimated variance components

From the cell means obtained from equation (3) by the above process the uncorrected sums of squares $T_a$, $T_b$, $T_{ab}$ and $T_f$ used in equations (2) can be calculated. Suppose we consider the first of these, namely

$$\hat{\sigma}_r^2 = \lambda_1 T_a + \lambda_2 T_b + \lambda_3 T_{ab} + \lambda_4 T_f + p_1 \hat{\sigma}_e^2 \ .$$

Since $\hat{\sigma}_e^2$ has a $x^2$-distribution we will rewrite this as

$$\hat{\sigma}_r^2 = \lambda_1 T_a + \lambda_2 T_b + \lambda_3 T_{ab} + \lambda_4 T_f + \theta_1 x^2 \ .$$

It is at once apparent that, for any particular n-pattern and set of values for the true variance components (and hence for any particular $\lambda$'s and $\theta$), the conditional distribution of $\hat{\sigma}_r^2$, given $T_a$, $T_b$, $T_{ab}$ and $T_f$, is $\theta_1 x^2$. i.e. the form of $p(\hat{\sigma}_r^2 \mid T_a, T_b, T_{ab}, T_f)$ is known.

Furthermore, as in the earlier calculations, the variance of $\hat{\sigma}_r^2$ has been obtained; and, since $\hat{\sigma}_r^2$ is unbiased its mean value is $\sigma_r^2$. Therefore, although the distribution of $\hat{\sigma}_r^2$ is unknown, it is possible to create empirically a set of contiguous intervals that encompass the complete range of possible values of $\hat{\sigma}_r^2$. For example if $V_r = \mathrm{Var}(\hat{\sigma}_r^2)$

$$
\begin{aligned}
I_1 &: -\infty < \hat{\sigma}_r^2 \leq \sigma_r^2 - 3.0V_r \\
I_2 &: \sigma_r^2 - 3.0V_r < \hat{\sigma}_r^2 \leq \sigma_r^2 - 2.3V_r \\
I_3 &: \sigma_r^2 - 2.3V_r < \hat{\sigma}_r^2 \leq \sigma_r^2 - 1.96V_r \\
&\vdots \\
&\text{etc.} \\
&\vdots \\
I_n &: \sigma_r^2 + 3.0V_r < \hat{\sigma}_r^2 < +\infty
\end{aligned}
$$

Decisions regarding the number of intervals and their magnitudes (the coefficients of $V_r$) would probably be based on the particular values of $\sigma_r^2$ and $V_r$ concerned. One would like to have the same number of intervals and the same coefficients of $V_r$ for all situations.

The known distribution

$$p(\hat{\sigma}_r^2 \mid T_a, T_b, T_{ab}, T_f)$$

is now used to calculate (from appropriate tables of the $\chi^2$-distribution) the probability

$$P_{r,k} = p(\hat{\sigma}_r^2 \text{ lies in } I_k \mid T_a, T_b, T_{ab}, T_f) \ .$$

This probability is calculated for each interval $I_k$; and it is calculated for every random sample drawn for any particular n-pattern and set of values of the true variance components. In each interval the mean probability is then calculated over all samples. Thus is established, by Monte Carlo methods, a histogram of the probability distribution of $\hat{\sigma}_r^2$ for a given n-pattern and a given set of values of the true variance components. This will be done for each component in the analysis, excluding $\sigma_e^2$, whose distribution is known. The whole process will be repeated for different n-patterns and different sets of values of the true variance components. If the intervals $I_1 \ldots I_n$ are taken reasonably small and sufficient in number, the histograms might be informative enough to display differences for different n-patterns and different sets of true variance components, if such differences exist.

Number of samples

As well as making decisions necessary for the setting up of the n-patterns and decisions on the sets of values to be used for the true variance components, one must also decide on the number of samples to take in any one case. 500-800 seems a nicely-sized range, but undoubtedly the amount of computer time needed for each sample will be a very determining factor.

## 3. Additional Topics

(a) Crump (1951) and Searle (1956) give expressions for the large sample variances of maximum likelihood estimators of the variance components in the 1-way classification. These could be calculated and compared with variances of estimators obtained from Henderson's Method 1. The maximum likelihood estimators themselves could possibly be obtained also, by iterative procedures.

Similar large sample variances are also available for the 2-way nested classification, (Searle, unpublished). The 2-way cross-classification appears to be intractable in this regard. The 3-way nested classification hasn't been tried.

(b) Henderson's other methods could be used - for both random and mixed models. Such work would require a lot of preliminary analysis - to find the sampling variances.

(c) Mahamunulu (1963) suggests a method of finding unbiased estimates of the sampling variances of estimated components. These could be calculated for each random sample and their distribution plotted - this may require sampling individual observations and not just cell means.

(d) Searle (1956) considers components of covariance in the 1-way classification. These could be considered too, perhaps.

(e) Bush and Anderson (1962) have developed matrix notation which merits close consideration, to see if it can simplify the necessary computing procedures.

## Acknowledgements

Many of the ideas presented above originated in discussions with colleagues, particularly with C. R. Henderson and D. S. Robson.

## References

Applied Research Laboratory, (1962). The generation of random samples from common statistical distributions. United States Steel, Technical Report, Project No. 25.17-016(1).

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of normal deviates, Ann. Math. Stat. 28, 610-611.

Bush, Norman and Anderson, R. L. (1962). Estimating variance components in a multi-way classification. Institute of Statistics, North Carolina State College, Mimeo Series No. 324.

Crump, S. C. (1951). The present status of variance component analysis. Biometrics 7, 1-16.

Henderson, C. R. (1953). Estimation of variance and covariance components. Biometrics 9, 226-252.

I.B.M. (1959). Random number generation and testing. Reference Manual C20-8011.

Mahamunulu, D. M. (1963). Sampling variances of the estimates of variance components in the unbalanced 3-way nested classification. Ann. Math. Stat. 34, 521-527.

Muller, M. E. (1959). A comparison of methods for generating normal deviates on digital computers. J. Assoc. Computing Machiner, 6, 376-383.

Searle, S. R. (1956). Matrix methods in variance and covariance components analysis. Ann. Math Stat. 27, 737-748.

Searle, S. R. (1958). Sampling variances of estimates of components of variance. Ann. Math. Stat. 29, 167-178.

Searle, S. R. (1961). Variance components in the unbalanced 2-way nested classification. Ann. Math. Stat. 32, 1161-1166.

Searle, S. R. (1963). Genetic studies of dairy production early in lactation. J. Dairy Sci. 47, (in press).

Table 1 Distribution of numbers of A.B. Daughters in herd-sire sub-classes in a study of 10,589 records in 654 herds by 119 sires, 1959-60. (i.e. numbers of paternal $\frac{1}{2}$ -sibs in the same herd)

| No. of daughters in a herd-sire subclass | No. of subclasses | % of all subclasses |
|---|---|---|
| 0 | 60,280 | 88.8745% |
| 1 | 5,472 | 8.0677 |
| 2 | 1,456 | 2.1467 |
| 3 | 406 | .5986 |
| 4 | 129 | .1901 |
| 5 | 54 | .0796 |
| 6 | 12 | .0177 |
| 7 | 11 | .0162 |
| 8 | 4 | .0059 |
| 9 | 1 | .0015 |
| 10 | 0 | |
| 11 | 1 | .0015 |

Totals 119 X 654 = 67,826 subclasses
7,546 occupied subclasses
10,589 daughters

Ratios: 11.1255% of subclasses were occupied
72.5152% of occupied subclasses had 1 observation
1.4032 observations per occupied subclass

Table 2 Distribution of Numbers of A.B. Daughters in 654 herds in New Zealand (1959-60)

| No. of A.B. daughters in a herd | No. of herds | % of herds | No. of A.B. daughters in a herd | No. of herds | % of herds |
|---|---|---|---|---|---|
| 1 | 3 | .5 | 31 | 6 | .9 |
| 2 | 11 | 1.7 | 32 | 6 | .9 |
| 3 | 31 | 4.7 | 33 | 3 | .5 |
| 4 | 30 | 4.6 | 34 | 10 | 1.5 |
| 5 | 33 | 5.0 | 35 | 7 | 1.1 |
| 6 | 29 | 4.4 | 36 | 4 | .6 |
| 7 | 30 | 4.6 | 37 | 6 | .9 |
| 8 | 26 | 4.0 | 38 | 3 | .5 |
| 9 | 31 | 4.7 | 39 | 3 | .5 |
| 10 | 33 | 5.0 | 40 | 2 | .3 |
| 11 | 27 | 4.1 | 41 | 3 | .5 |
| 12 | 31 | 4.7 | 42 | 2 | .3 |
| 13 | 22 | 3.4 | 43 | 4 | .6 |
| 14 | 25 | 3.8 | 44 | 3 | .5 |
| 15 | 22 | 3.4 | 45 | 1 | .1 |
| 16 | 20 | 3.1 | 46 | 1 | .1 |
| 17 | 18 | 2.7 | 47 | 0 | - |
| 18 | 15 | 2.3 | 48 | 1 | .1 |
| 19 | 20 | 3.1 | 49 | 2 | .3 |
| 20 | 18 | 2.7 | 50 | 2 | .3 |
| 21 | 13 | 2.0 | 51 | 1 | .1 |
| 22 | 15 | 2.3 | 52 | 2 | .3 |
| 23 | 14 | 2.1 | 53 | 1 | .1 |
| 24 | 10 | 1.5 | 54 | 1 | .1 |
| 25 | 9 | 1.4 | 55 | 1 | .1 |
| 26 | 10 | 1.5 | 56 | 3 | .5 |
| 27 | 5 | .8 | 58 | 2 | .3 |
| 28 | 7 | 1.1 | 59 | 1 | .1 |
| 29 | 9 | 1.4 | 75 | 1 | .1 |
| 30 | 5 | .8 | | | |

Totals 654 herds,
10,589 daughters