A STUDY OF NEW OBJECTIVE YIELD PROCEDURES FOR FILBERTS

by

William H. Wigton

Research and Development Branch
Standards and Research Division
Statistical Reporting Service

March 1970

Table of Contents

A STUDY OF NEW OBJECTIVE YIELD PROCEDURES FOR FILBERTS

by

William H. Wigton

I. Introduction

Objective estimates for filberts in Washington and Oregon were made from 1955 to 1964. They were dropped in 1965 because the objective estimates were not accurate. This survey was resumed with matched industry and SRS funds in 1968. For 1968 and 1969, filbert production that actually reached processors was considerably more than the direct expansion estimates of biological production from the objective counts.

The present study was designed to test ways of improving the definition of sampling units, the sample allocation, and techniques for reducing the counting errors in an effort to improve objective yield estimates. The study was based on data collected from six filbert blocks in Oregon. These blocks represented a variety of orchard types.

II. Objectives

The objectives of this study were to determine better sampling methods and survey procedures for collecting objective yield information. Estimating procedures were compared for sampling efficiency. Counting errors were measured for terminal limbs. Photography of bare trees was evaluated for possible use as a sampling frame.

III. Summary

1. The sum of the primary 1/ limb CSA's 2/ for a tree is highly correlated with the estimated total number of nut clusters from the tree. The measurements of primary CSA's are inexpensive to obtain; therefore, the double sampling scheme 3/ for selecting trees is feasible and efficient.

1/ Primary scaffolds or primary sampling units were major limb divisions emerging from the main trunk.

2/ Cross-sectional area.

3/ The term double sampling is used when an inexpensive measurement is made on a large sample. Then a subsample of these selected for a more expensive measurement and the mean of the small sample is adjusted for the difference between it and the mean of the large sample (see Section V).

2. Primaries should be selected using equal probabilities. The size of the primary limb should be used in the estimation process in a regression estimator.

3. Terminal sample units should be selected using equal probabilities provided a reasonable range (.8 to 2.5) is placed on their size. Expansions to the primary limb stage should be made disregarding the size of the terminal limb.

4. The optimum sample allocation within a block is:

(a) Three trees
(b) One primary limb on each tree
(c) Two terminal limbs per primary

The total time required to do this, including the tree selection, is between two and one-half and three hours.

5. All clusters of nuts on the selected terminal limbs should be picked and bagged. An independent quality count should follow a few days after the first stripping, to determine whether the proper limb was stripped and whether any nut clusters were missed. The proper size of the subsample would depend on percent of undercounts and funds.

6. Bare tree photography of sample trees can be used for selecting the primary and terminal limbs. These pictures would be extremely useful in the quality check work and for rotating sample limbs in succeeding years.

7. The use of photography to count nut clusters on trees a month before harvest has not been fully evaluated. Preliminary indications raise serious doubts about its feasibility. Counts for slides analyzed show that only a small percentage of nut clusters are visible on the photograph.

IV. Survey Methodology

A. Sample Selection

Eight filbert blocks were selected by the Oregon State Statistical Office (SSO) for this research study. Rough sketches of the blocks were available where each tree was represented by a square on a piece of graph paper. These sketches indicate four things: (1) The number of rows of trees in the block, (2) approximate number of trees in each row, (3) approximate number of trees for the entire block, and (4) exactly where the blocks were located in relation to barns, fields, houses and roads which bordered the blocks.

Permission to enter the blocks was obtained from the operator of the orchard. These blocks were then visited in March when the trees were bare. A random sample of three or four rows was systematically selected. From these rows, eight or nine trees within each row were again systematically selected. This assured that the trees were fairly evenly spaced throughout the block. From these trees (about 30) the trunk and primary limb measurements were obtained by a special tape from which square inches could be read directly.

From previous work on other tree crops it is known that the sum of the primary CSA's is more highly correlated with total yield for a single tree than the one measurement of trunk CSA. Therefore, the CSA of the primary limbs on the trees were added to a tree total and these sums were arrayed by size. Then three trees were systematically selected from the array. The subsampled trees were located again, flagged with engineering tape, and photographed from two sides. A stereo camera was used so that the three dimensional effect could be utilized to separate and identify limbs.

In the office, primary and terminal limbs 4/ were defined as: (1) A primary SU 5/ (limb) is a major branch of the tree that has limitations on its maximum and minimum size and (2) a terminal SU (limb) is a branch with a CSA measurement between .8 and 2.5 inches.

Itek negative prints were made of each stereo slide. From the stereo slide the trees were then broken into primary limbs. Two of these primaries per tree were selected to be subsampled. Each of the two primaries was broken down into terminal limbs and two per primary limb were chosen for objective counts. All sample units were marked on the Itek print, but no limbs were measured.

B. Field Procedures for Collecting Counts

In August 1969 the eight blocks were again visited by Oregon and R&D personnel. However, two blocks were not used--one because of excessive deer damage and the other because of time restrictions. The analysis that follows is based on the remaining six blocks.

The required times to complete all job phases were recorded. The trees previously selected were located again. All the primary limbs, as marked on the Itek prints, were measured. In most trees one or more limbs were not within the defined range. If the limbs were too large, they were broken into two or more primary limbs. If the primary limb did not have two

---

4/ Sample unit.
5/ A complete set of definitions may be found in Appendix A.

acceptable terminal limbs it was combined with another so that the combination was within the defined range for primary limbs. All sizes were recorded. When either of these types of changes was made in the field it meant that a new selection of two primary limbs had to be made. Regardless of where the primary limbs were selected the primaries were partitioned on the photograph into terminal limbs. These units were checked to make sure they were within the defined range. If any limbs did not comply with the definitions, the appropriate changes were made before the two were selected. Again, all terminal limb sizes were recorded; however, selections were made using equal probability.

The nut clusters on the selected limbs were then counted. The two primaries were assigned at random to the two counters. The man from the Research and Development Branch used the method of partitioning the limb and counting by sections, while the man from the Oregon SSO counted the limbs by the procedure outlined in the Interviewer's Manual. 6/

Any small limbs (those less than .8 square inches) that did not have a probability of selection at some stage, were treated as "path units" and cluster counts made for them. (Copies of the field procedures and forms that were used are shown in Appendix A.)

After each man had his counts on both terminal limbs on the primary the men changed primary scaffolds and selected one terminal limb from the pair for stripping. Every cluster on the limb was then picked and the limb was checked again to make sure it was completely stripped. The nut clusters were put into plastic bags which were identified by block, tree, and limb on tree. This procedure was done on three trees in each of the six blocks. The bags were taken to the state laboratory, and the clusters were broken up so that the individual nuts could be counted.

V. Model Tested

Double sampling model requires making measurements on many trees and selecting a subsample of those measured for objective counts and measurements. For double sampling to be effective a second variable is needed that is highly correlated with the variable being estimated and inexpensive to obtain. For filberts, nut clusters is the variable being estimated. Tree size is a useful auxiliary variable to use in estimation because it can be estimated by using many characteristics. The characteristics that seem most efficient are trunk CSA and the sum of primary limb CSA's because they take only a few minutes to obtain for each tree.

6/ Interviewer's Manual. Filbert Objective Yield Survey, Oregon.

In order to evaluate the advantage of double sampling, the double sampling model was used:

$$\hat{M}_i = \bar{Y}_i + b\ (\bar{X}_{il} - \bar{X}_{is})$$

Where: $\hat{M}_i$ is the improved estimate of nut clusters for ith block.

$\bar{Y}_i$ is the average direct expansion estimate for the three trees in the ith block.

b is the slope of the regression line of X_{ij}, the sum of primary CSA's on the jth tree in the ith block, and Y_{ij} the total number of nut clusters on the same tree. $(Y_{ij} = a_i + bX_{ij})$

$\bar{X}_{is}$ is the average sum of primary CSA's for the small sample of trees for which counts were made.

$\bar{X}_{il}$ is the average sum of primary CSA's for the large sample of trees.

The associated variance function is:

$$\text{Within block variance} = \underbrace{\frac{S_t^2(r^2)}{n'} + \frac{S_t^2(1-r^2)}{n}}_{\text{between tree variance}} + \underbrace{\frac{S_p^2}{nm} + \frac{S_{ter}^2}{nml}}_{\text{within tree variance}}$$

S_t^2 = variance component between trees.

S_p^2 = variance component between nut clusters on primary scaffolds within trees.

S_{ter}^2 = variance component between nut clusters on the terminal sample units within primary scaffolds.

r^2 = correlation coefficient squared between total nut clusters and the covariate measurement; i.e., trunk CSA or sum of primary scaffolds CSA's.

n' = the number of trees for which measurements were obtained.

n = the number of trees in the subsample selected for objective counts.

m = the number of primaries per tree.

t = the number of terminal sample units per primary limb.

The amount of actual gain in terms of reduced variance for this model depends on three things:

1. The amount of correlation between total nut clusters on a tree and tree size.

2. The magnitude of the between tree variance component compared to the magnitude of the within tree variance.

3. The number of observations of the large and small samples.

One gets a better estimate of the regression slope ($\hat{b}$) if the selected trees have a wide range of sizes. This is because the variance of $\hat{b}$ is

$\frac{S_e^2}{\Sigma x^2}$ (the larger the Σx^2 the smaller the variance). For this reason the trees were selected systematically from an array of trees by sizes.

A. Analysis Trees

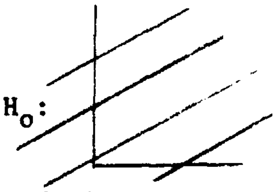
The first phase was to determine a suitable regression model for pooling the data from the different blocks. A sequential test procedure was used for this purpose, starting with the most complex model and proceeding to the least complex model. A detailed explanation will be given only for the first model of this type. The regression coefficients of the model $\hat{Y}_{ij} = a_i + b_i X_{ij}$ were tested where $\hat{Y}_{ij}$ is estimated total nut clusters for the jth tree in the ith block, and X_{ij} is the trunk CSA (or sum of the primary CSA's) for the jth tree in the ith block. The a_i and b_i are the within block regression coefficients for the ith block.

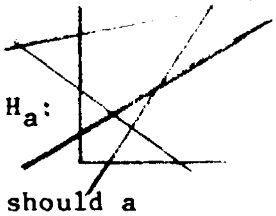
Table 1 is an analysis of variance (AOV) summary which tests various hypotheses about the suitability of regression lines when combining data gathered in the different blocks. The test is terminated with the first significant F value. The AOV table, which is read starting at the bottom, tests the following:

1. Can an average within block slope be used for all pooled data, or is a different slope and intercept necessary for each block?

   H_o: $\hat{Y}_{ij} = a_i + b\ X_{ij}$

   H_a: $\hat{Y}_{ij} = a_i + b_i\ X_{ij}$

   H_o: 

   H_a: 

2. Can one intercept (or mean) and slope be used or should a common slope but separate intercept be used for each block?

   H_o: $\hat{Y}_{ij} = a + b\ X_{ij}$

   H_a: $\hat{Y}_{ij} = a_i + b\ X_{ij}$

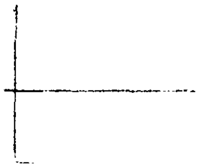
   H_o: 

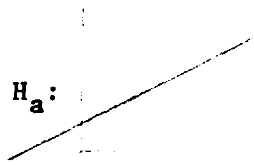
   H_a: 

3. Is a regression equation useful or would the mean, $\bar{Y}$, be appropriate; i.e., is b = o?

   H_o: $\hat{Y}_{ij} = \bar{Y}_i$

   H_a: $\hat{Y}_{ij} = a + b\ X_{ij}$

   H_o: 

   H_a: 

Once these questions are answered, the basic estimating model is established.

The top part of Table 1 is a standard AOV table for the expanded cluster counts. This top section shows the partitioned sums of squares which will be used to compute the correlation coefficient.

The first F value of 1.52 is not significant, thus H_o: $\hat{Y}_{ij} = a_i + b\ X_{ij}$ is accepted and the next test is considered. The second F value is significant; therefore, H_a: $Y_{ij} = a_i + b\ X_{ij}$ is the proper model.

An average within block slope may be used for all blocks. This slope predicts ($\hat{Y}$) for a unit change in the trunk CSA (X). The model ($\hat{Y}_i = a_i + bx_{ij}$) can be changed to the double sampling model ($\hat{M}_i = \bar{Y}_i + b\ (\bar{X}_{i1} - \bar{X}_{is}$) by observing that $a_i = \bar{Y}_i - b\bar{x}_{is}$, and $\bar{X}_{i1}$ is the large sample value for the covariate.

Table 1.- An analysis of variance testing various hypotheses about the suitability of regression lines. 1/

Source of variation	Degrees of freedom	Sums of squares	Mean square	F-test	Hypotheses
Between groups....................	5	17,829,392	3,565,878	2/ 8.62	H_o: $\sigma_B^2 = 0$
Within groups.....................	12	4,962,781	413,565		H_1: $\sigma_B^2 \neq 0$
Total corrected sums of squares...	17	22,792,174			
Regression (a, b).................	1	1,566,472	1,566,472		H_o: $\hat{Y}_{ij} = Y$
Error 1...........................	16	21,225,701	1,326,606		H_a: $\hat{Y}_{ij} = a+bX_{ij}$
Regression ($a_1 \ldots a_6$, b)...........	5	16,948,263	3,389,653	2/ 8.72	H_o: $\hat{Y}_{ij} = a+bX_{ij}$
Error 2...........................	11	4,277,438	388,858		H_a: $\hat{Y}_{ij} = a_i+bX_{ij}$
Regression ($a_1 \ldots a_6$, $b_1 \ldots b_6$).....	5	2,391,979	478,396	1.52	H_o: $\hat{Y}_{ij} = a_i+bX_{ij}$
Error 3...........................	6	1,885,459	314,243		H_a: $\hat{Y}_{ij} = a_i+b_iX_{ij}$

1/ X = trunk cross-sectional area, Y = estimated total of nut clusters.
2/ Indicates significance at 1 percent level.

Table 2.- An analysis of variance testing various hypotheses about the suitability of regression lines. 1/

Source of variation	Degrees of freedom	Sums of squares	Mean square	F-test	Hypotheses
Between groups	5	17,604,936	3,520,987	2/ 8.51	H_o: $\sigma_B^2 = 0$
Within groups	12	4,962,782	413,565		H_1: $\sigma_B^2 \neq 0$
Total corrected sums of squares	17	22,567,718			
Regression (a, b)	1	2,740,612	2,740,612		H_o: $\hat{Y}_i = \bar{Y}$
Error 1	16	19,827,106	1,239,194		H_1: $\hat{Y}_i = a+bx$
Regression ($a_1 \ldots a_6$, b)	5	17,360,221	3,472,044	2/15.8	H_o: $\hat{Y}_i = a+bx$
Error 2	11	2,466,885	224,262		H_1: $\hat{Y}_i = a_i+bx$
Regression ($a_1 \ldots a_6$, $b_1 \ldots b_6$)	5	745,640	149,128	.52	H_o: $\hat{Y}_i = a_i+bx$
Error 3	6	1,721,245	286,874		H_1: $\hat{Y}_i = a_i+b_ix$

1/ X = sum of primary scaffold cross-sectional area, Y = estimated total of nut clusters.
2/ Indicates significance at 1 percent level.

The same tests were made for the regression coefficients where X_{ij} is the sum of the primary CSA's. These tests are displayed in Table 2.

The first F value of .52 is not significant. The null hypotheses ($Y_i = a_i + bX_{ij}$) is accepted and the next test is considered. The F value

of 15.8 is highly significant; therefore, the testing stops. The model for grouping this data is $Y_{ij} = a_i + b\ X_{ij}$. It is the same as before.

A model for combining the data has been established, so that now the correlation coefficients could be calculated. The within-block correlations had to be computed by summing up sums of squares adjusted for the block means, and using these values to figure the correlation in the usual manner; i.e.:

$$r^2 = \frac{(\Sigma\Sigma xy)^2}{(\Sigma\Sigma x^2)\ (\Sigma\Sigma y^2)}$$

These sums of squares are displayed in Tables 3 and 4 for trunk CSA (X variable in Table 3), and sum of primary CSA's (X variable in Table 4.)

Two blocks were fairly homogeneous with respect to the X variable. These blocks can be identified by observing the adjusted Σx^2 column in Tables 3 and 4.

The correlations were computed twice--once using all blocks and once removing the two homogeneous blocks. Table 5 is a summary of these correlations with associated degrees of freedom.

The correlation coefficient for the sum of primary scaffolds and estimated total nut clusters is highly significant. Neither of the correlations for trunk CSA is significantly different from Zero at the .05 level.

The cost of obtaining these measurements in terms of time should be broken into two parts: (a) The time it takes to walk from one tree to another, and (b) the time it takes to make the various measurements at the tree. Time required to go from one tree to another would be the same for either variable, trunk CSA or sum of primary CSA's, while time at the tree for obtaining the sum of the primaries is about three times as long as obtaining the trunk CSA. However, the time required for both measurements is only four minutes per tree for one person. These measurements need not be redone each year. Initial measurements could be used for a period of four or more years. We recommend use of the sum of primary scaffolds, since it is more highly correlated with clusters of nuts per tree. The increased cost is nominal when it is spread over the total number of years that the measurements could be used.

Table 3.- Adjusted sums of squares for trunk C.S.A. and estimated total nut clusters. 1/

Block	k	Degrees of freedom n-1	Adjusted $\Sigma_x2 = A_i$	Adjusted $\Sigma_{xy} = B_i$	Adjusted $\Sigma_y2 = C_i$
253	1	2	1,580.0867	23,706.6667	492,370.67
2/ 271	2	2	439.0067	-23,726.1700	1,554,868.70
277	3	2	940.6067	23,217.2000	651,534.00
274	4	2	1,968.0067	7,356.1667	36,016.67
325	5	2	1,458.1667	31,654.6667	1,708,181.00
2/ 328	6	2	145.4467	4,695.8667	519,810.67
Within groups............		$\Sigma n_i - k$	$\Sigma A_i = A_w$	$\Sigma B_i = B_w$	$\Sigma C_i = C_w$
		12	6,531.3202	66,904.4004	4,962,781.71
Among groups.............		k-1	$A_m = A_t - A_w$	$B_m = B_t - B_w$	$C_m = C_t - C_w$
		5	10,099.9020	94,503.0196	17,829,392.09
Total sums of squares....		$\Sigma n_i - 1$	A_t	B_t	C_t
		17	16,631.2220	161,407.4200	22,792,173.80

1/ X = trunk C.S.A., Y = estimated total nut clusters.
2/ These blocks are considered homogeneous.

Table 4.- Adjusted sums of squares for the sum of the primary scaffold C.S.A. and estimated total nut clusters. 1/

Block	k	Degrees of freedom n-1	Adjusted $\Sigma_x 2 = A_i$	Adjusted $\Sigma_{xy} = B_i$	Adjusted $\Sigma_y 2 = C_i$
253	1	2	2,247.1667	29,451.6667	492,370.67
2/ 271	2	2	497.3267	-9,262.3700	1,554,868.70
277	3	2	1,690.9600	32,976.6000	651,534.00
274	4	2	3,539.3867	5,768.3334	36,016.67
325	5	2	1,592.8867	51,548.5700	1,708,181.00
2/ 328	6	2	562.7267	14,279.1330	519,810.67
Within groups............		$\Sigma n_i - k$	$\Sigma A_i = A_w$	$\Sigma B_i = B_w$	$\Sigma C_i = C_w$
		12	10,130.4535	159,011.1997	4,962,781.71
Among groups............		k-1	$A_m = A_t - A_w$	$B_m = B_t - B_w$	$C_m = C_t - C_w$
		5	8,412.5885	66,420.1003	17,604,935.79
Total sums of squares....		$\Sigma n_i - 1$	A_t	B_t	C_t
		17	18,543.0420	225,431.3000	22,567,717.50

1/ X = sum of primary scaffold C.S.A., Y = estimated total nut clusters.
2/ These blocks are considered homogeneous.

Table 5.- Within block correlations between trunk C.S.A. and sum of primary C.S.A. and estimated total nut clusters.

Measurements	Correlation for six blocks	Degrees of freedom	Correlation with two homogeneous blocks removed	Degrees of freedom
Trunk C.S.A. vs. estimated total nut clusters............	.3716	11	.6551	7
Sum of primary scaffolds vs. estimated total nut clusters..	.7091	11	.9514	7

B. Analysis Within Tree Clusters

Two estimators were considered for expanding clusters on the primary scaffolds, and on the terminal sample units to tree totals:

1. Simple unbiased expansions.

2. Using size as an auxiliary variable in a regression or unbiased ratio expansion.

For ratio estimates to be more effective, a correlation must be sufficiently high; i.e., r is greater than 1/2 $\frac{(S_x)}{\bar{X}} / \frac{(S_y)}{\bar{Y}}$, but even with perfect correlation 7/ the ratio of $\frac{b^2}{a^2} \frac{(1-f)}{n}$ must be greater than $\frac{\bar{X}^2 \; V \; (1/\bar{X})}{S_x^2}$, or the estimates will be less precise where the a and b come from the regression equation.

If the correlation is large enough, then the next criterion must be met. This second criterion is less binding since one can reduce the Y intercept by a simple transformation. For example, if the correlation is high and the slope is large, but the intercept is also large (Diagram 1), a simple transformation of the X-variable can reduce the Y intercept to zero (Diagram 2).

7/ Des Raj, Sampling Theory, page 92.

Diagram 1

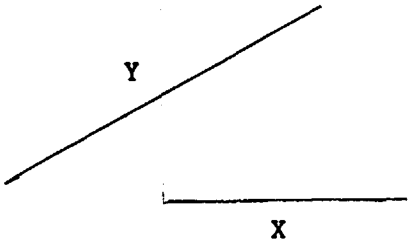
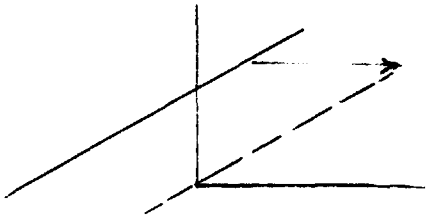


Diagram 2



The regression estimator is not restricted by the value of a single Y intercept if a within block model is used.

1. Primary Limbs

The ranges specified for the size of the primary limbs required each to have at least two terminal sample units and that they not be larger than 1/4 of the trunk CSA. Thus, a wide range of sizes is possible for primary scaffolds.

To evaluate which estimation procedure is more efficient for primary limbs, r was compared with $1/2 \frac{(S_x)}{\bar{X}} / \frac{(S_Y)}{\bar{Y}}$. In order to compute the correlation coefficient, it is necessary to see how the data should be combined. Table 6 shows these tests for combining data.

The first hypothesis tested was:

H_o: $\hat{Y}_{ij} = a_i + b\ X_{ij}$ against the alternative

H_a: $\hat{Y}_{ij} = a_i + b_i\ X_{ij}$

Can one average slope be used for the data in the six blocks or is a separate slope needed in each block? The F value (.80) is not significant. One average slope can be used; the next test is then considered:

H_o: $\hat{Y}_{ij} = a + b\ X_{ij}$

H_a: $\hat{Y}_{ij} = a_i + b\ X_{ij}$

Table 6.- Analysis of variance on the regression equations. 1/

Source of variation	Degrees of freedom	Sums of squares	Mean square	F-test	Hypotheses
Between groups....................	5	383,028	76,605	7.01	H_o: $\sigma_B^2 = 0$
Within groups.....................	30	327,732	10,924		H_1: $\sigma_B^2 \neq 0$
Total corrected sums of squares...	35	710,760	20,307		
Regression (a, b).................	1	63,747	63,747		H_o: $\hat{Y} = \bar{Y}$
Error 1...........................	34	647,013	19,030		H_1: $\hat{Y} = a + bX$
Regression ($a_1 \ldots a_6$, b)...........	5	431,619	86,324	2/11.62	H_1: $\hat{Y}_i = a + bX$
Error 2...........................	29	215,394	7,427		H_1: $\hat{Y}_i = a_i + bX$
Regression ($a_1 \ldots a_6$, $b_1 \ldots b_6$).....	5	30,794	6,159	.80	H_o: $\hat{Y}_i = a_i + bX$
Error 3...........................	24	184,000	7,692		H_1: $\hat{Y}_i = a_i + b_iX$

1/ X = cross-sectional area of the primary scaffold, Y = estimated total nut clusters on the primary scaffold.
2/ Indicates significance at 1 percent level.

Can one average intercept be used for the data in the six blocks or is a separate intercept needed in each block? The F value (11.62) is highly significant. Therefore, the alternative hypothesis is accepted and the testing procedure terminated. The proper model for utilizing data from the different blocks is $\hat{Y}_{ij} = a_i + b X_{ij}$, where b is the average within block slope for all data and a different intercept (a_i) for each ith block must be computed. Y is total clusters on primary limb in ith block on jth primary and X_{ij} is size of the jth primary in the ith block.

The correlation coefficient (assuming one average slope) can be computed by: (1) Subtracting error 2 SS from within groups of Y, (2) dividing the difference by within groups of Y, and (3) taking the square root. Table 7 shows these sums of squares.

Table 7.- The within blocks SS used to compute correlation coefficient. 1/

Source of variation	Sums of squares
Within groups of Y................	327,732
Error 2..........................	215,394
Regression assuming one b.........	112,338

$r^2 = \frac{112,338}{327,732} = .343$ $\quad r = \sqrt{.343} = .586 \quad (S_x / \bar{X}) = .608$

$(S_y / \bar{Y}) = .521 \quad 1/2\ (S_x / \bar{X}) / (S_y / \bar{Y}) = .583$

The correlation coefficient is slightly larger than $1/2\ (S_x / \bar{X}) / (S_y / \bar{Y})$. It should be pointed out that this relationship is based on an approximation of the mean square error of the classical ratio estimation. The inequality is approximate and if the correlation is high and the slope large (as this case) the size information may still be helpful. The correlation computed was between X_{ij} (the size of the jth primary in the ith block) and the $\bar{Y}_{ij}$ (estimated number of clusters on the same limb).

If the correlation were run between X_{ij} and Y_{ij} (the actual count of nut clusters for that primary) the correlation would be increased.

The model which was obtained in Table 6 for combining the data at the primary level suggests that the regression model would be the proper way to use the size information to improve the estimates. The regression model does not assume that the intercepts for each block are zero.

The second inequality necessary for the ratio estimator to be efficient involves the slope and the intercepts:

$$\frac{b^2}{a^2} \frac{(1-f)}{n} > \frac{\overline{X}^2 V\ (1/\overline{X})}{S_x^2}$$

This inequality was looked at in a different form.

$$\frac{b^2\ S_x^{\ 2}}{\overline{X}^2\ (V\ (1/\overline{x})\ )} \frac{(1-f)}{n} > a_i^{\ 2}$$

After substituting the proper values, it turns out that $|a_i|$ must be less than 64. The intercept for each block was computed:

$$a_1 = 42,\ a_2 = 125,\ a_3 = -7,\ a_4 = 111,\ a_5 = 327,\ a_6 = 12.$$

In three blocks, it would be helpful to use the ratio estimator and in the other three, the ratio estimator would be less efficient than the simple direct expansion. Furthermore, one cannot change the intercept by a single linear transformation because the intercepts vary so widely. For this reason, a within block regression estimator is recommended, using the following model:

$\hat{Y}_i = \overline{Y}_i + b\ (\overline{X}_{i1} - \overline{X}_{is}) = \overline{Y}_i + b\overline{X}_{i1} - bX_{is}$ where $\overline{Y}_i - b\overline{X}_{is}$ is the block intercept (a_i).

$\hat{Y}_i$ is estimated average number of total clusters on a tree in ith block. $\overline{Y}_i$ is average direct expansion estimate for the ith block. b is the regression slope, $\overline{X}_{i1}$ is the average primary size for the block, $\overline{X}_{is}$ is the average size of the primaries that were sampled.

Since the correlation squared is (.34), and the slope is significant, the primary limb size data will probably reduce the primary variance component by about 1/3. However, data is needed on more blocks to evaluate the actual amount of reduced variances caused by using CSA of primary limbs in the estimation process. The recommendation is to collect the additional information but to select the primaries with equal probabilities. This procedure would enable variances to be computed both ways.

2. Terminal Limbs Within Primary Limbs

The primary SU's must be broken into terminal SU's. This unit was defined as any limb with a CSA between .8 and 2.5 square inches. It averaged 50 nut clusters and took approximately 13 minutes to count. Two estimation schemes were studied. Equal probability selection with:

1. Expansion by reciprocal of probability.

2. Expansion using size as an auxiliary variable in a ratio or regression estimate.

To determine which method of estimation is more efficient r was compared to $1/2\ (S_x/\bar{X})/(S_y/\bar{Y})$. The results of Table 8 were used to obtain the proper model for combining the data.

The first test, starting at the bottom, tests whether or not one average slope can be used in all blocks:

$$H_o:\ \hat{Y}_{ij} = a_i\ b\ X_{ij}$$

$$H_a:\ \hat{Y}_{ij} = a_i + b_i X_{ij}$$

The F value is .62 which is not significant.

The F value of the second test:

$$H_o:\ \hat{Y}_{ij} = a + b\ X_{ij}$$

$$H_a:\ \hat{Y}_{ij} = a_i + b\ X_{ij}$$

is 15.46 which is highly significant.

The model (H_a: $\hat{Y}_{ij} - a_i + b\ X_{ij}$) is the accepted model for combining data for the terminal stage.

The correlation is computed by subtracting error 2 SS from the within group SS and dividing by the within group SS.

Table 8.- An analysis of variance on the regression equations. 1/

Source of variation	Degrees of freedom	Sums of squares	Mean square	F-test	Hypotheses
Between groups	5	44,851	8,970	2/ 14.95	H_o: $G_B^2 = 0$
Within groups	64	38,406	600		H_1: $G_B^2 \neq 0$
Total corrected sums of squares	69	83,257			
Regression (a, b)	1	1,157	1,157		H_o: $\hat{Y}_i = \bar{Y}$
Error 1	68	82,100	1,207		H_1: $\hat{Y}_i = a + bX$
Regression ($a_1 \ldots a_6$, b)	5	45,222	9,044	2/ 15.46	H_o: $\hat{Y}_i = a + bX$
Error 2	63	36,878	585		H_1: $\hat{Y}_i = a_i + bX$
Regression ($a_1 \ldots a_6$, $b_1 \ldots b_6$)	5	1,864	373	.62	H_o: $\hat{Y}_i = a_i + bX$
Error 3	58	35,014	604		H_1: $\hat{Y}_i = a_i + b_iX$

1/ X = cross-sectional area of the terminal sample unit, Y = nut clusters on this unit.
2/ Indicates significance at 1 percent level.

Table 9.- The within blocks SS used to compute correlation coefficient.

Source of variation	Sums of squares
Within groups of Y...............	38,406
Error 2...........................	36,878
Regression assuming one b.........	1,528

$r^2 = \frac{1528}{38406} = .04$ $\quad\quad 1/2 \left(\frac{S_x}{\bar{x}} / \frac{S_y}{\bar{y}}\right) = .34$

$r = .2$[illegible]

The first criterion necessary for size to be used in the estimation procedure using a ration estimator is not met. Neither the ratio nor the regression estimation scheme which uses the terminal size would reduce the variance because the r^2 is very low. Therefore, if the terminal SU's are restricted in size from .8 to 2.5 square inches then the use of simple unbiased estimates are more efficient than to have sizes enter into the estimation process.

VI. Optimize Sample Allocation

A. Optimum Number of Trees, Primary Limbs, and Terminal Limbs

Two sample allocations have been optimized:

1. Optimum values for trees (n), primaries within trees (m), and terminals within primaries (t).

2. The optimum ratio of trees measured to trees counted was computed.

Both optimizations were done assuming that equal probability selection and estimation would be used at all stages. The variance components were computed from sample data selected in this manner. However, the estimating model for the average tree within kth block is:

$$\bar{Y}_{k\cdot} = \frac{1}{n} \sum_{i=1}^{n} \frac{(r_{ki})}{m} \sum_{j=1}^{m} \frac{(S_{kij})}{t} \sum_{w=1}^{t} X_{kijw}$$

where X_{kijw} is number of filbert clusters on wth limb, the jth primary, on the ith tree in the 1th block.

t is the number of terminal sample units selected.

S_{kij} is number of terminals on the jth primary in the ith tree in the kth block.

m is the number of primary sample units selected.

r_{ki} is the number of primary sample units on the ith tree in the kth block.

n is the number of trees per block.

Its associated variance formula is:

$$\text{Total variance} = \frac{S_B^2}{k} + \frac{S_T^2}{kn} + \left(\frac{\bar{M}-m}{\bar{M}}\right) \frac{S^2{}_P}{kmn} + \left(\frac{\bar{T}-t}{\bar{T}}\right) \frac{S^2{}_{Ter}}{knmt}$$

and the appropriate cost function is: Total cost = (k) C_B + (kn) C_T + (knm) C_p + (knmt) C_{Ter}

where:

k is number of blocks in sample.

S_B^2 is variance component between blocks.

C_B is cost of going from block to block (or block to home).

$S^2{}_T$ is variance component between trees.

C_T is cost of going from tree to tree within a block and breaking the tree into primary units.

$S^2{}_P$ is variance component between primaries.

C_P is cost of selecting one primary and breaking it into terminal sample units.

S^2_{Ter} is variance component between terminals within primaries.

C_{Ter} is cost of counting and selecting one terminal.

$\bar{M}$ is average number of primaries on a tree = 5.89.

$\bar{T}$ is average number of terminals on a primary = 5.

In order to simplify, average finite correction factors were computed and the proper variance components were reduced accordingly. The sampling fraction was very small at the block and tree level so that the finite corrections could be ignored.

According to Cochran 8/, the optimum values for t, m, and n are:

$$t = \sqrt{\frac{C_P \; S^{2'}_{Ter}}{C_{Ter} \; S^{2'}_P}} \qquad m = \sqrt{\frac{C_T \; S^{2'}_P}{C_P \; S^{2'}_T}} \qquad n = \sqrt{\frac{C_B \; S^2_T}{C_T \; S^2_B}}$$

The numerical values which were substituted are found in Table 10.

Table 10.- Summary of costs and variance components adjusted for average finite correction factors for the four stages. 1/

Source	Costs in minutes	Variance components
Blocks	150	118,113 = S^2_B
Trees	18	115,519 = S^2_T
Primaries	9	91,334 = $S^{2'}_P$ 2/
Terminal sample units	16	554,293 = $S^{2'}_{Ter}$ 2/

1/ This table summarizes major findings from Tables 12, 13, 14, 15 and 16 in Appendix B.
2/ These components have been reduced by average finite correction factors.

8/ Snedecor and Cochran, "Statistical Methods," pages 632-3.

The optimum values rounded to integers are:

$n = 3$

$m = 1$

$t = 2$

We may now obtain the optimum ratio of trees measured to trees counted.

B. Optimum Ratio

To optimize the ratio $\frac{n^1}{n}$, again variance and cost function are necessary. This time within-block functions are needed:

$$\text{Within-block variance} = \frac{S^2_T}{n} + \frac{S^2_{P'}}{nm} + \frac{S^2_{Ter'}}{nmt}$$

This must be revised to include double sampling at the tree level: Within-block double sampling variance =

$$\frac{S^2_T\ (r^2)}{n'} + \frac{S^2_T\ (1-r^2)}{n} + \frac{S^2_{P'}}{nm} + \frac{S^2_{Ter'}}{nmt}$$

and a within-block double sampling cost =

$$n'\ C'_T + n\ C_T + nm\ C_P + nmt\ C_{Ter}$$

where C'_T is cost of measuring a tree, four minutes per tree, but could be used for four years so that one minute average per year was used. n' is the number of trees selected at random to measure. r^2 is the correlation coefficient squared between the estimated quantity (total nut clusters) and the auxiliary variable, the measure of tree size. In this study we have recommended sum of primaries as the covariate and assumed an r^2 of .7.

The optimum ratio can now be found by:

1. Forming the product of the variance and cost functions.

2. Differentiating with respect to n' and n.

3. Solving for each and forming the ratio.

4. Substituting the appropriate values.

The ratio before substitution is:

$$\frac{n'}{n} = \sqrt{\frac{S^2{}_T\ (r^2)\ (C_T + m\ C_P + (mt)\ C_{Ter}}{C_T'\ (S^2{}_T\ (1-r^2) + \frac{S^2{}_{P}'}{m} + \frac{S^2{}_{Ter}'}{mt}}}$$

which is 3.4 after substitution. Since three trees per block should be selected for counts, 10.2 is the optimum real number to measure for the double sample. It is recommended that 12 trees be selected for measurements because 12 is a multiple of three and a rotation system could be worked out with 12 trees.

VII. Counting Errors

Since the counts were resumed in 1968, the filbert limbs have been undercounted. Table 11 shows limb counts, direct expansion estimates, strip counts, and revised estimates for the terminal SU's in the research data.

When the number of nut clusters missed (stripped counts minus on tree cluster counts) are plotted against stripped counts, the graphs indicate that a proportional relationship exists. A fitted line has a positive slope and goes approximately through the origin.

This would indicate that a percentage of undercount could be applied to a limb count. To obtain a percentage of undercount, a subsample of terminal limbs must be selected at random and stripped. Each enumerator should have some limbs in the subsample.

Counting errors can be reduced so that they are negligible. Since new terminal SU's are much smaller than the limbs in the old sample, the recommendation is to strip the limbs completely for the count. The required time for a quality check would be mostly made up of time to get to the block and locate the tree.

Table 11.- Comparison of nut counts on limbs without stripping to stripped counts. (Oregon filberts)

Block	Tree	Expanded cluster counts for four limbs	On tree cluster counts for two limbs		Stripped counts for same two limbs		Stripped counts		Revised estimate for four limbs
			Limb 1	Limb 2	Limb 1	Limb 2	Limb 1	Limb 2	
328	44-28	1,065	6	38	8	42	1.33	1.11	1,231
	9- 5	285	25	4	31	4	1.24	1.00	351
	18-12	1,048	61	69	74	80	1.21	1.16	1,237
325	31-10	4,230	66	13	67	15	1.02	1.15	4,484
	21-20	2,399	129	117	154	157	1.19	1.34	3,023
	21-10	2,476	54	153	68	167	1.26	1.09	2,773
271	1-16	1,295	19	44	20	48	1.05	1.09	1,373
	6-11	1.109	56	23	88	37	1.57	1.61	1,752
	6-21	2,480	41	15	51	15	1.24	1.00	3,054
253	38- 7	876	68	63	74	65	1.09	1.03	934
	1/ 6- 3	31	24	--	31	--	1.29	--	40
	22- 3	754	42	75	53	76	1.26	1.01	860
277	8- 6	1,486	33	72	38	72	1.15	1.00	1,590
	26-10	433	33	65	37	66	1.12	1.02	450
	14-17	897	7	62	7	69	1.00	1.11	969
274	16-13	1,084	20	13	18	8	.90	.62	845
	12-13	790	17	16	16	21	.94	1.31	790
	22-13	437	7	11	8	18	1.14	1.64	594

1/ Very small tree.

VIII. A Summary of Recommended Within-Block Selection and Estimation Procedures

The following is a summary of the recommended within-block sampling allocation based on 1969 results.

A. Selection of Trees

1. A sample of 12 trees should be selected at random for each block for the purpose of obtaining primary limb measurements. Make the measurements and array the trees by size.

2. A subsample of three trees from the 12 should be systematically selected for the purpose of making objective counts.

3. A regression estimator would be used to adjust the subsample mean of three for differences from the larger group mean of 12 trees.

B. Selection of Limbs Within the Tree

1. In each of the three trees selected for objective counts, select one primary limb using equal probabilities.

2. Record sizes for all the primary limbs.

3. Break down the primary limb into terminal limbs.

4. Select two terminal limbs for the purpose of making counts.

5. Strip the two limbs of all nuts.

6. Estimate tree total by using a regression model where the direct expansion estimate is adjusted for size differences in the primary limbs selected and the average primary size of those measured.

IX. Estimated Time Requirements for Recommended Sample Allocation

These times are based on results of Appendix B Tables 12 and 13. Field time the first year for selecting trees and obtaining measurements, subsampling and selecting two terminal SU's on each tree would be three hours and 45 minutes.

		Minutes
1.	Select 12 trees at random and obtain measurements.....	48
2.	Select a subsample of three trees and walk back to the trees..	24
3.	Check the primary sampling units to make sure they are in the defined range and select one per tree - ten minutes per tree..................................	30
4.	Divide the primary into terminal sample units - nine minutes per tree.................................	27
5.	Select and strip two limbs by clusters - 32 minutes per tree..................................	96
	Total time..	225

Since the measurements on the 12 trees would be used for a period of four years, the years following the first year would require less than three hours:

		Minutes
1.	Locate the trees and select a primary sampling unit - 18 minutes per tree...........................	54
2.	Divide the primary into terminal SU's - nine minutes per tree.................................	27
3.	Select, strip, and bag the two terminal units per tree - 32 minutes per tree...........................	96
	Total time..	177

The new recommended procedure requires more time than the average survey field time in 1968 (142 minutes) or 1969 (112 minutes).

The required enumerator time in the block can be cut by using bare tree photography, or sample selection procedures based on a previous visit to the block ahead of the main survey.

X. Use of Bare Tree Photography as a Sampling Frame

This year in the six research blocks, bare tree stereo photographs were used as a sampling frame to select sampling units. The limbs were selected in the office. When the selected limbs were located in the field, many were not within the proper size limitations. Consequently, changes had to be made in the field. Some of the reasons for this were:

1. No measurements were made in the photographs.

2. Some thinning was done after the pictures were taken.

3. Enough care was not taken in the selection process.

4. The camera was not focusing properly.

If the stereo slides with Itek prints are to be used in an operational survey, some changes in 1969 field procedures would be advisable. To begin with, much time was wasted in the field because all the limbs were measured to check sizes. If any did not fall in the defined ranges for that stage of selection, the total stage was partitioned again and new random selections were made. This would not have been necessary if some measurements had been made on the Itek prints.

Another method which was used to equalize the size of the limbs is the following: Before taking the stereo slides, flag (with colored tape) the primary SU's and flag one limb of .8 square inches and another of 2.5 square inches which are easily visible in the photograph. Then, when partitioning the primary SU, a visual comparison is possible; i.e., if a small limb is perhaps too small to be called a terminal SU, visually compare it with the flagged limb of .8 in the photograph. If the limb in question is smaller than the flagged one, you know it is path and too small to be called a terminal SU.

This system worked very well in Michigan on cherry trees.9/ The average coefficient of variation was .409 by using the latter method. The .409 is less than the average coefficient of variation obtained in 1969 where the limbs were measured and changed if they were too large or small.

The primary SU's are already visibly marked in the photograph so that an acceptable primary SU will be selected. The measurements for the terminal SU's are not needed for the estimating process.

The terminal SU's are selected in the office and marked on the Itek prints. The Itek prints are given to the enumerators who, with the help of the slides and prints, locate the terminal SU's, strip the limbs, check the limb again and put all clusters in a bag which is identified for each tree and block.

9/ Michigan Cherry Research Project, Fred Vogel.

A followup quality check is very simple and inexpensive. The supervisor takes the Itek print, visits the block, locates the terminal limb, checks to make sure the proper limbs were stripped. The actual check of the limb would take about two minutes since all he must do is to run his hand quickly over the limb perhaps finding one or two clusters.

The expected counting error by this procedure would be about two percent.

The costs involved if bare tree photography is to be used would be:

Travel costs (figure three blocks per day)
Agricultural Statistician salary 1/3 day

Slides, film and processing..........................$2.10
30 cents per slide, seven slides per block

Itek prints..$2.10
30 cents per print, seven prints

Time in block in March:

	Minutes
Time to select 12 trees...................	48
Select three trees and walk back to them...	24
Time to mark limbs and take stereo pictures (20 minutes per tree)	60
Time to leave block.......................	12
Total................................	144

Time in the block in August to strip limbs would be reduced to:

	Minutes
Time to walk to trees and locate limbs..... (18 minutes per tree)	54
Time to strip limbs....................... (32 minutes per tree)	96
Total................................	150

The same photographs would be used for more than one year.

XI. The Best Linear Estimator for Combining Data 10/

It is possible that several techniques will be utilized the next few years for obtaining filbert limb data. For example, in some blocks, the method of counting one primary per tree will still be used. In other blocks the new procedure may be used and in still other blocks both procedures may be used, but on different trees.

The purpose of this section is to recommend a scheme which would combine all data in one final estimate and at the same time show which scheme is most efficient. The best weight function 11/ can combine estimates whose expected values are the same; i.e., $E(t_i) = u$, with covariance terms not necessarily equal to zero; i.e., $\text{cov}(t_i, t_i) = a_{ii}$. Weights may be found such that:

1. The sum of the weights equal one.

   W_i $(i = 1, 2, \ldots P)$. $\sum_{i=1}^{P} W_i = 1$

2. The expected value to the linear combination is the population parameter.

   $E(\Sigma W_i t_i) = u.$

3. Variance of the linear combination is minimum.

To find these weights, one forms the variance-covariance matrix (A) of all the estimators. If three estimates are being combined, the variance-covariance matrix will be a three by three. The inverse matrix (A^{-1}) must be found and all of the weights can be obtained from this inverse.

The weight for the first estimate, t_1, is the sum of the elements in the first column divided by the sum of all elements in A^{-1}. The sum of the elements in the second column divided by the sum of all elements in A^{-1} is W_2, etc. The variance of the linear combination is the reciprocal of sum of the elements in A^{-1}. If the estimates are independent the procedure is the same as using the reciprocals of their variances as weights.

10/ Des Raj, Sampling Theory, pages 16-17.

11/ A program for the Remote Access Terminal (RAX) which computes the weights is available. See Appendix page RAX Program S160WT (Best Linear Estimator).

The following two estimators were combined by the best linear weight function. This is presented to demonstrate the procedures.

Estimator 2 - the simple direct expansions ~~adjusted for differences in primary sizes~~. The double sampling of trees was not used. The model on page 21 is estimator 2.

Estimator 1 - the simple direct expansion is adjusted for differences in tree size. The model is $M_{ki} = Y_{ki} + b\ (X_k - X_{ki})$

M_{ki} = adjusted estimate of ith tree in kth block.

Y_{ki} = estimator 1.

b = slope of within-block regression coefficient where Y is estimated total clusters, and X is sum of cross-sectional areas of primary limbs.

X_i = average of primary CSA's in large sample in kth block.

X_{ki} = sum of primary CSA on ith tree in kth block.

Since two estimators will be combined, the variance-covariance matrix (A) will be a two by two.

$$A = \begin{matrix} 868311 & 858652 \\ 858652 & 1022379 \end{matrix}$$

$$A^{-1} = \begin{matrix} .000006795 & -.000005707 \\ -.000005707 & .000005771 \end{matrix}$$

Column 1 total = .000001088

Column 2 total = .000000064

Grand total = .000001152

$$\text{Weight 1} = \frac{\text{Column 1 total}}{\text{Grand total}} = .944 \qquad \text{Weight 2} = \frac{\text{Column 2 total}}{\text{Grand total}} = .056$$

Mean estimator 1 = 1195 = M_1. Mean estimator 2 = 1299 = M_2.

The best single estimator is t_1 which can be seen by inspecting the diagonal terms of the A matrix. The best linear estimator is $W_1M_1 + W_2M_2 = 1200$. Its associated variance is

$$\frac{1}{\text{Grand total}} = 867773.$$

In this example M_1 is a better estimator, therefore it gets the most weight. Its variance alone was 868,311. The best linear estimator did not improve this much because most of the weight went to the one estimate. (Both estimates came from the same counts on the same trees.)

XII. Future Work

In order to verify that the size of the selected primary will reduce the variance, the recommendation is to compute variances two ways:

1. For simple unbiased estimate.

2. For a regression estimator which adjusts for differences in sizes between the selected primary and the mean for the tree.

The data from the first year would provide enough evidence to answer the question.

The problem which needs further study is that of converting the number of nuts to weights. The recommendation is to first estimate the total number of nuts per tree. Then, in a separate step, convert the nuts to size groups and weights. A procedure for estimating nut sizes in August and a conversion procedure should be verified by data at harvest.

XIII. Appendixes

Definitions of Sampling Units and Path Sections

Primary Limb:

A primary limb is a major branch of a tree with limitations on its maximum and minimum size. It is less than 1/4 the total CSA measurement of the sum of the "scaffold" limbs. Also a primary limb must have at least 2 "terminal units." A terminal unit is defined below. In most cases, a primary limb will correspond to a scaffold limb as defined in the Oregon Interviews Manual; i.e., a major branch of the tree.

For a bush type tree, limbs satisfying the definition of a terminal unit must be combined in groups of two or more units to be a primary limb.

Trunk Path:

Branches or "suckers" too small to be called terminals that originate from the ground or off the main trunk below the primary limbs.

Primary Path:

Branches or small twigs (on a section of a primary) too small to be classified as terminals. The path count excludes any terminal limbs which emerge from it.

Terminal Unit:

A branch with a CSA measurement of between .8 inch and 2.4 inches hopefully. Two limbs can be combined to form one terminal unit provided each is between .5 and .7 and they are close together. Terminal units arising from the main trunk are to be assigned to primaries.

Field Procedures

(Oregon Filberts)

1. Record time you leave office in morning.

2. Record time arrived at block.

3. Record time arrived at first tree.

4. Identify and tape (using red engineering tape) all primary limbs as marked in Itek print. Mark tape with magic marker.

    a. Record primary limbs with CSA measurements.

    b. Record trunk measurements.

    c. Record time when finished.

5. Start with selected primary limb most near starting corner of block.

   a. Map this section completely using red tape and blue tape (yellow may have to be substituted) for terminal units.

   b. Record all measurements.

   c. Break down terminals that are too large and combine up to <u>two</u> that are too small. Terminal units should have a CSA of 1.6 sq. inches; however, limbs between .8 and 2.4 CSA are acceptable.

   d. Record changes on field form and Itek print.

T.U.	CSA		CSA
i.e. 2	4	2a	______
		2b	______

   e. Randomly select two terminal units without replacement.

   f. Record time on field form after both primaries are mapped.

6. Enumerator records time on his form and counts two terminals using "his procedure" then records time.

7. Statistician counts second primary (two terminals).

8. People change primaries and strip 1 terminal unit on each primary.

9. Recheck stripped limbs for missed clusters, bag nuts and times.

10. Move to next tree and begin with step one.

Oregon Filbert Survey (Statistician) Field Form

Block _______

Tree _______ Date _______

Time left last block or office _______

Time arrived at block _______ Trunk C.S.A. _______

Time arrived at first tree _______

Primary	CSA	Primary	CSA
_______	_______	_______	_______
_______	_______	_______	_______
_______	_______	_______	_______
_______	_______	_______	_______
_______	_______	_______	_______

Time when finished with this phase _______ Expansion factor at this point _______

Selected Primary _______ Selected Primary _______

Terminal Unit	CSA	Changes	CSA	Terminal Unit	CSA
_______	_______	_______	_______	_______	_______
_______	_______	_______	_______	_______	_______
_______	_______	_______	_______	_______	_______
_______	_______	_______	_______	_______	_______
_______	_______	_______	_______	_______	_______
_______	_______	_______	_______	_______	_______
_______	_______	_______	_______	_______	_______
_______	_______	_______	_______	_______	_______
_______	_______	_______	_______	_______	_______

Time when finished first primary _______ Expansion factor for first _______

Time when finished second primary _______ Expansion factor for second _______

Counts on Back

Time begin first terminal count ________ Time begin second count ________

Time end first count ________ Time end second terminal count ________

Prim.	Term.	Sect.	Counts	Prim.	Term.	Sect.	Counts

Time when tree is finished ____________

Appendix B - Variances and Costs

Table 12.--Table showing estimated average time required to complete certain job phases in man hours

Source of time	Average time required
Average time to measure one tree	3 minutes
*Average time required to arrive at a block leaving from the office	40 minutes
Average time required to arrive at the first tree in the block	8 minutes
Average time to locate selected scaffolds, measure scaffolds, and map and select terminal limbs or sample units/primary scaffold	38 minutes
**Average time to count one terminal	13 minutes
Average time to strip and bag clusters, complete forms and pack equipment	25 minutes

*Only six blocks were used to get this average.
**Time required to strip limb is equal to or less than time required to count it.

Table 13.--Estimated times used in optimizing sample allocation (Some of the times were changed for optimization)

Source of time	Average time required	
Average time required to drive from block to block or block to home plus mileage..........................	*40 minutes	150 minutes
Average time to locate first tree, carry all equipment, and measure primary scaffolds..................	18 minutes	
Average time to select one primary and break it into all possible sample units and record terminal sample unit sizes.................................	25 minutes	9 minutes
Average time to select and count one terminal sample unit..	16 minutes	

*The between block cost of 40 minutes does not consider mileage. This between block cost was changed to 150 minutes. It is made up of mileage cost (38 miles/block times 10¢) 3.80 G.S. 3 time is 96 minutes plus driving time 50 minutes between block cost 150. This mileage reflects costs between home of blocks.

**Since the CSA of the terminal limbs will not need to be collected, it would take much less than 25 minutes. We feel 9 minutes would be adequate.

Table 14.--Analysis of variance table for estimated total nut clusters on trees. (This AOV is for the six blocks)

Source	DF	Sums of squares	Mean square	F-test	Variance components
Blocks	5	49374889	9874978	6.236	*710335
Trees	12	19001474	1583456	1.391	115467
Primary Scaffolds	18	20487739	1138208	1.232	110256
Sample units	34	31409929	923821		923821
Total	69	120274031	1743102		

* This between-block component was based on 5 degrees of freedom. Another between-block component was calculated from the data from the 327 blocks. See Table 15.

Table 15.--Analysis of variance for estimated total nut clusters for survey data collected in 1969

Source	DF	Sum of squares	Mean square	Variance Components
Blocks	326	215302432	660437	158260
Trees	654	121420544	185658	185658
Total	980	336722944	343595	

The survey data was collected from one sample cluster sample of three trees in each block. Therefore, the between-block component of variance is overstated because the data from the sample clusters includes between cluster variation within the block as well as between block variation. No measure of the between sample cluster component is available from the 327 blocks. The total sums of squares on the survey data of 1969 was divided by using the between primary and between tree components for the research blocks and then finding a new between block component. Table 16 gives a more accurate estimate of the between block variance component.

Table 16.--Variance components, mean squares and sums of squares to compute new between block variance component

Source	DF	Sums of squares	Mean squares	Variance components
Blocks	326	189100102	580062	118113
Trees	654	147622842	225723	*225723
Total	980	336722944		

*This component comes from Table 14. The variance component for trees (115467) and the variance component for primaries (110256) are added.

Appendix C

RAX Program S16ØWT (Best Linear Unbiased Estimator)
(F. B. Warren, January 1970)

Problem

Given a set of K unbiased estimators of a parameter Y, t_i, i = 1, 2,..., k, k-7 and the k x k variance-covariance matrix (A) of these k estimators is given by cov (t_i t_j).

Compute $\underline{w} = \frac{\underline{e}\ A^{-1}}{\underline{e}\ A^{-1}\underline{e}'}$ where the w_i can be used to compute a weighted average of the Y_i which will be the minimum variance estimator of Y, and $\underline{e}$ is a kX1 vector of 1's.

Note: This differs from the standard multiple regression estimator approach in that elements a_{ij} of A may come from different sources.

Solution

1. Read k, and the k x k A matrix of variance of covariances.
2. Compute A-1
3. Sum of the elements in the individual rows of A^{-1} eA^{-1}
4. Sum eA^{-1} $eA^{-1}e'$
5. Compute $\underline{w}'_i = eA^{-1}$ / (eA-1e')
6. $w_i = \frac{w_i'}{\Sigma w'}$

Operation

/ INPUT
/ INCLUDE S16ØWT (This program is stored on cards and must be saved prior to use.)
/ DATA
Enter data cards (More than one problem set can be entered sequentially in the same data deck.)
/ END RUN

Data Preparation

(k+1) cards must be prepared for each problem set. Data cards will be prepared as follows:

Card Number	Format	Description
1	I1	Number of linear estimators to be considered.
2, 3,...,k+1	(10X, 7F10.3)	The first, second, ..., kth rows of the variance-covariance matrix for the k variables.