

New Methods for Filbert Objective Yield Estimation

By William H. Wigton and William E. Kibler

Filbert estimating techniques can be improved by using refined procedures for selecting sample limbs and counting nut clusters. These procedures can reduce survey cost 25 percent and improve sampling and nonsampling errors. Counting nut clusters for two terminal limbs (4 percent of an average tree) by stripping them from limbs reduced counting errors considerably, compared with on-the-limb counts for primary limbs (15 percent of an average tree). The total cross-sectional area of primary limbs is inexpensive to obtain and can be used efficiently in a double sampling survey design.

Key words: Sampling; estimation; statistical methodology; fruit and nut counting.

Filbert production estimates for Washington and Oregon were made from 1955 to 1964 using both objective yield procedures¹ and data reported by growers (3).² The objective estimates were discontinued for economic reasons until 1968, when the demand for more complete and accurate information on quality and quantity of the crop increased. This paper discusses some work that has been done to increase the accuracy of objective yield estimates by improving (1) the definition of sampling units, (2) sample allocation, (3) estimating procedures, and (4) field counting procedures. The work described has applications for other fruit and nut crops where objective yield procedures have been or are being considered.

Sample Selection

Six filbert blocks (orchards) were used in the study. Rough sketches of the blocks were made with each tree represented by a square on graph paper. The sketches also indicated (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) location of the blocks in relation to barns, fields, houses, and roads bordering the blocks (figure 1).

A systematic sample of three or four rows and eight

or nine trees in each row was selected in each block for the total study, using random starts. This assured a uniform distribution of sample trees throughout the block as shown in figure 1. The trunk and primary limb³ measurements (cross-sectional areas or CSA's) of these trees were measured by using a special tape which is read directly in square inches.

Previous work (1) on other tree crops indicates that the sum of the primary CSA's for a given tree is more highly correlated with total yield than the one measurement of trunk CSA. Therefore, the sample trees were arrayed by the sums of the CSA's of their primary limbs. A subsample of three trees was systematically selected from this array as shown in table 1. Detailed counts and measurements were made for these three trees. The subsampled trees were flagged with engineering tape and photographed from two opposite sides during dormancy. A stereo camera was used so the three-dimensional effect could be used to identify limbs. The stereo slides were used to partition the trees into sampling units, first by identifying the primary limbs. Two randomly selected primaries per tree were further subdivided into terminal limbs.⁴ All sample units (terminal limbs) were identified on photographs. Two terminals from each primary were chosen as sample units for making counts of nut clusters. Individual nuts cannot be identified until nuts are mature and hulls open so the individual nuts drop out. A cluster generally contains about four nuts but can have as few as one or as many as eight nuts.

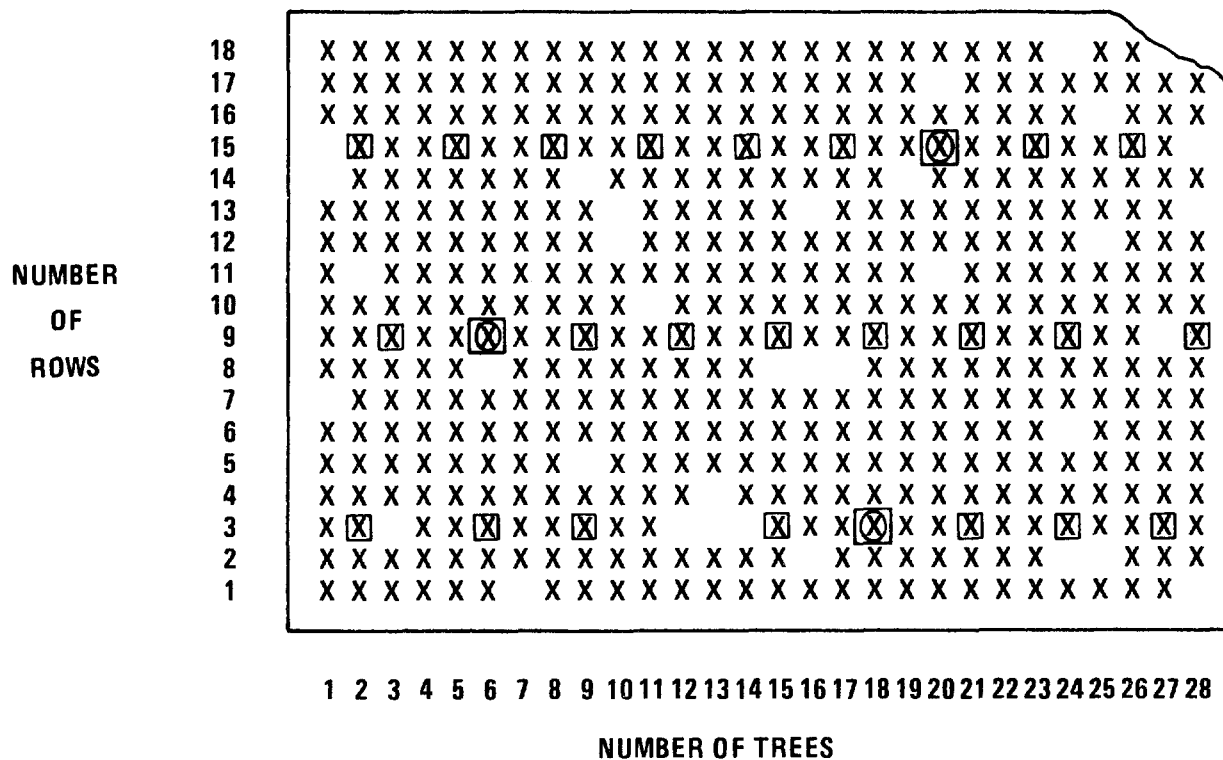
¹ Estimation procedures based on actual plant or fruit characteristics measured or counted from randomly selected plots or limbs.

² Italic numbers in parentheses indicate items in the References, p. 46.

³ Primary limbs or scaffolds are major limb divisions emerging from the main trunk (figure 2).

⁴ Small limbs emerging from the primary limbs used as sample units for counting nuts (figure 2).

**ORCHARD SKETCH SHOWING SAMPLE TREES WITHIN BLOCK:
ABOUT 475 TREES IN BLOCK OF 18 ROWS, 28 TREES PER ROW**



☒ SAMPLE TREES USED FOR PRIMARY CSA MEASUREMENTS
⊗ SUBSAMPLE OF TREES USED FOR MAKING CLUSTER COUNTS ON TERMINAL LIMBS

Figure 1

Table 1.—Sum of primary cross-sectional areas for sample trees shown in figure 1 arrayed with subsampled trees identified

Row and tree	Sum of primary CSA's <i>Square inches</i>	Row and tree	Sum of primary CSA's <i>Square inches</i>
Row 9 Tree 12	93.4	Row 3 Tree 15	81.0
Row 15 Tree 5	89.7	Row 3 Tree 21	80.4
Row 3 Tree 24	88.2	Row 15 Tree 11	78.9
Row 15 Tree 20	^a 86.4	Row 3 Tree 9	78.7
Row 9 Tree 3	85.8	Row 9 Tree 24	77.6
Row 3 Tree 6	85.7	Row 15 Tree 26	77.6
Row 3 Tree 27	84.9	Row 15 Tree 8	76.2
Row 15 Tree 2	83.8	Row 3 Tree 2	76.0
Row 9 Tree 18	83.3	Row 3 Tree 18	^a 75.1
Row 3 Tree 15	82.6	Row 15 Tree 23	74.5
Row 15 Tree 14	82.2	Row 9 Tree 9	72.8
Row 15 Tree 17	81.6	Row 9 Tree 28	70.1
Row 9 Tree 6	^a 81.4	Row 9 Tree 21	67.8

^aSubsample of trees for making cluster counts on terminals.

Field Procedures for Counting Clusters

In August, the selected trees were located again and all the primary limbs, identified on the photographs, were measured. Additional restrictions were placed on the size of the "primary limb" to help control variability. Its CSA could not be more than one-fourth of the sum of the CSA's of all primary limbs and it had to have at least two terminal limbs. One or more primary limbs on most trees were not within this range. If limbs were too large, they were divided into two or more primary limbs. Primary limbs without two acceptable terminal limbs were combined with another primary so the combination was within the defined range. This required a new selection of primary sample limbs and a partitioning of them into terminal limbs in the field. The CSA's of all terminal limbs on the selected primaries were

TRUNK AND LIMB STRUCTURE FOR A TYPICAL FRUIT OR NUT TREE

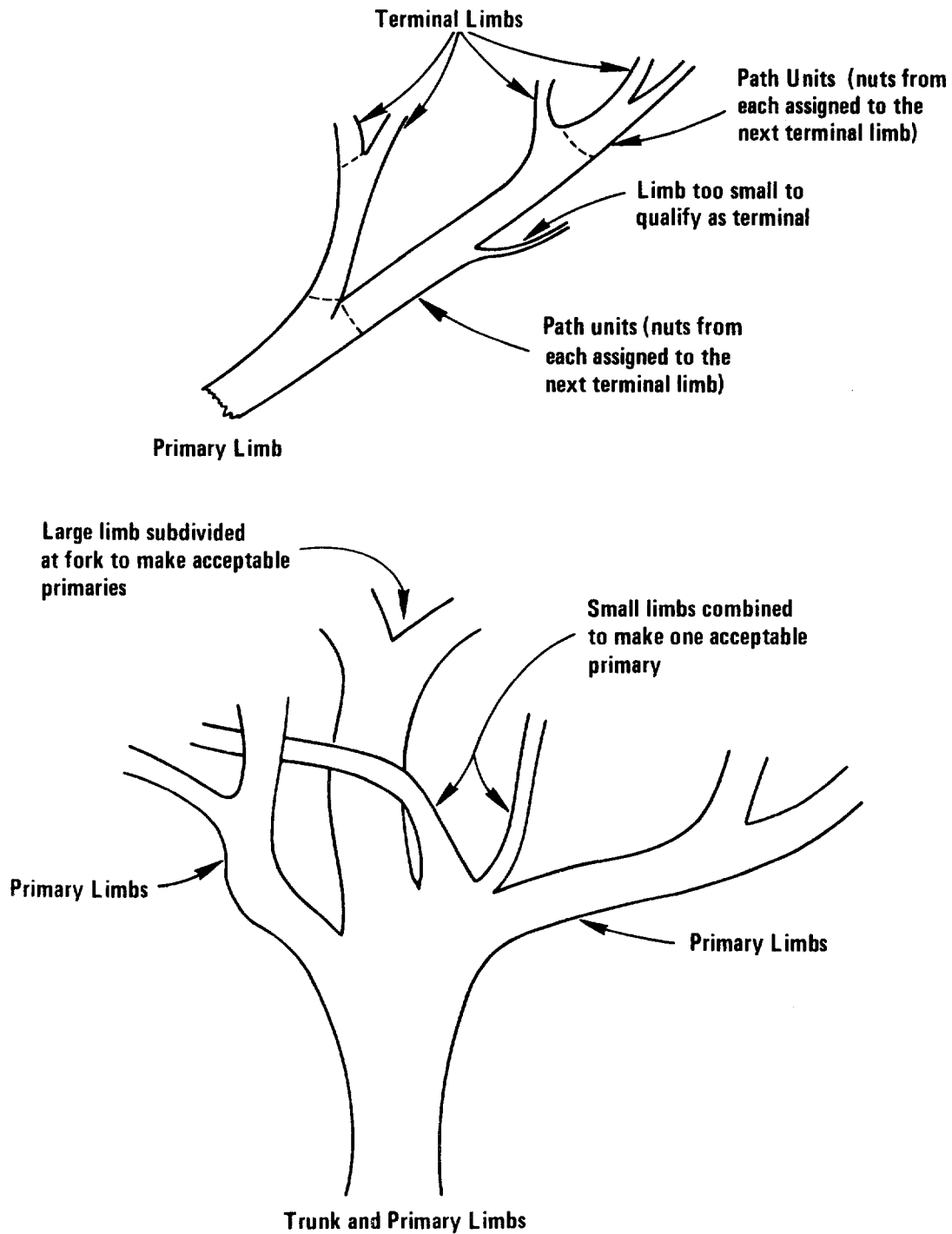


Figure 2

recorded. However, actual selections were made with equal probabilities. Any small limbs on a selected primary which had a CSA less than 0.8 square inch were treated as "path units" and the clusters were assigned to the closest terminal.

The nut clusters on the selected limbs were counted by two men. Sample limbs were assigned at random to the two counters. One counter used the method of partitioning the limb and counting by sections. A second man counted the nuts by starting at the base of the sample limb, counting outward, and recording one on a hand counter after each 15th cluster. After each completed his counts, they changed primary limbs and picked every cluster from one of the terminal limbs. This provided a quality check for the two counting procedures. The nut clusters were put in plastic bags, identified by block, tree, and limb, and sent to the State laboratory. Here the clusters were divided so individual nuts could be counted. The time required to complete each phase of the field work was recorded.

Estimating Models Evaluated

Two different types of models were considered in addition to the simple unbiased (direct expansion) estimate. In general, both the regression and ratio estimators use a double sampling approach. For a double sampling design to be effective, the related characteristic (auxiliary variable) must be highly correlated with the value being estimated and relatively inexpensive to obtain (compared with the variable under study). The double sampling designs were evaluated on two levels: (1) To estimate the number of nut clusters in a tree, and (2) to estimate the number of nut clusters on a primary. At each level, two possible covariates were studied—sum of primary CSA's and trunk CSA at the block level, and primary and terminal CSA's at the tree level.

Block Estimating Model—Regression

The model for the regression estimator is:

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

where

\hat{Y}'_i is the new estimate of nut clusters per tree for i th block

\bar{Y}_i is the average number of nut clusters estimated per tree based on the three trees subsampled in the i th block

b is the slope of the regression line of Y_{ij} , the total number of nut clusters on the same tree, on X_{ij} , the sum of primary limb CSA's (or trunk CSA) on the j th tree in the i th block

\bar{X}_{is} is the average CSA of all primary limbs (summed) for the subsample of trees for which nut cluster counts were made

\bar{X}_{il} is the average CSA for all primary limbs (summed) for the large sample of trees.

The associated variance function is:

Within-block variance =

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

where

S_t^2 = variance component between trees

S_p^2 = variance component between numbers of nut clusters on primary limbs within trees

S_{ter}^2 = variance component between numbers of nut clusters on terminal units within primary limbs

r^2 = coefficient of determination between total nut clusters and the covariate measure; i.e., trunk CSA or sum of primary limb CSA's

n' = number of trees for which CSA measurements were obtained

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

m = number of primaries selected per tree

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

The amount of actual gain in terms of reduced variance for this model depends on (1) the degree of correlation between total nut clusters on a tree and tree size, (2) the magnitude of the between-tree nut count variance compared with the magnitude of the within-tree

nut count variance, and (3) the number of observations for the large and small samples. The estimate of the regression slope b is better if the selected trees vary considerably in size (such as sum of primary CSA's). This is because the variance of b is $S_e^2/\Sigma x^2$ where S_e^2 is the mean square deviation from regression. The larger the Σx^2 , the smaller the variance of b . Thus, trees for the detail study were selected systematically from a list of trees arrayed by sizes (sum of primary CSA's).

The first step in testing the suitability of a regression model is to determine whether tree data from different blocks can be pooled. A sequential test procedure, starting with the most complex model and proceeding to the least complex model, was used.

This procedure is an analysis of variance (AOV) which tests a sequence of hypotheses about the suitability of combining data from different blocks in computing the regression coefficients. The following sequence of hypotheses is terminated with the first significant F value.

(1) Can an average within-block slope be used for all pooled data, or is a different slope and intercept necessary for each block (figure 3)?

$$H_0: \hat{Y}_{ij} = a_i + bX_{ij}$$

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

(2) Can one intercept (or mean) and slope be used or should a common slope, but separate intercept, be used for each block (figure 4)?

$$H_0: \hat{Y}_{ij} = a + bX_{ij}$$

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

(3) Is a regression equation useful or would the mean, \bar{Y} , be appropriate; i.e., is $b = 0$ (figure 5)?

$$H_0: \hat{Y}_{ij} = \bar{Y}_i$$

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

The basic estimating model is established by answering these questions.

The top part of table 2 is a standard AOV table for the estimated number of cluster counts. This top section shows the partitioned sums of squares used to compute the correlation coefficient. In testing the sequence of hypotheses, one starts at the bottom of table 1 and works up. The first F -value (1.52) is not significant; thus, $H_0: \hat{Y}_{ij} = a_i + bX_{ij}$ is not rejected and the next test is considered. The second F -value is significant; there-

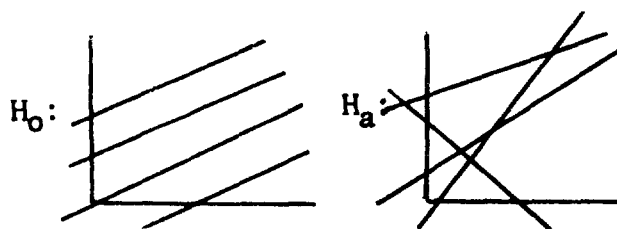


Figure 3

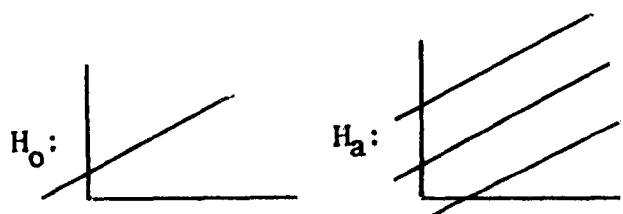


Figure 4

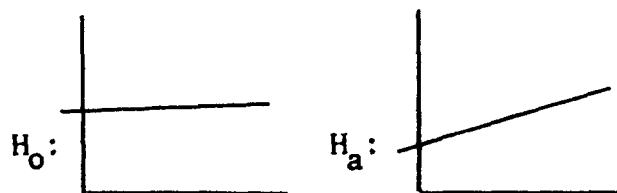


Figure 5

fore, $H_a: \hat{Y}_{ij} = a_i + bX_{ij}$ is the model indicated by the data.

An average within-block slope may be used for the trunk CSA's for all blocks. This slope predicts (\hat{Y}_{ij}) for a unit change in the trunk CSA (X_{ij}). The regression model ($\hat{Y}_{ij} = a_i + bX_{ij}$) is changed to the double sampling model ($\hat{Y}_i = \bar{Y}_i + b(\bar{X}_{il} - \bar{X}_{is})$) (2) by observing that $a_i = \bar{Y}_i - b\bar{X}_{is}$, where \bar{X}_{il} is the large sample value for the covariate and \bar{X}_{is} is the value for the small sample.

Similar results were obtained using the sum of the primary CSA's rather than the trunk CSA as the independent variable. Again, the first F -value (0.52) is not significant. The null hypothesis ($\hat{Y}_i = a_i + bX_{ij}$) is accepted and the next test is considered. The next F value (15.48) is highly significant and the testing stops. The model for grouping these data is $\hat{Y}_{ij} = a_i + bX_{ij}$, the same as for the trunk CSA.

After establishing the model for combining the data, correlation coefficients were computed. The within-block correlations were computed by adding the sums of squares adjusted for the block means, and using these values to figure the correlation in the usual manner; i.e.:

Table 2.—Analysis of variance testing various hypotheses about the suitability of regression lines ^a

Source of variation	Degrees of freedom	Sums of squares	Mean square	F-test	Hypothesis
Between blocks	5	17,829,392	3,565,878	^b 8.62	$H_0: \bar{Y}_i - \bar{Y}_k = 0$
Within blocks	12	4,962,781	413,565	-----	$H_a: \bar{Y}_i - \bar{Y}_k \neq 0$
Total corrected sums of squares	17	22,792,174	-----	-----	
Regression (a, b)	1	1,566,472	1,566,472	-----	$H_0: \hat{Y}_{ij} = \bar{Y}$
Error 1	16	21,225,701	1,326,606	-----	$H_a: \hat{Y}_{ij} = a + bX_{ij}$
Regression (a ₁ ...a ₆ , b)	5	16,948,263	3,389,653	^b 8.72	$H_0: \hat{Y}_{ij} = a + bX_{ij}$
Error 2	11	4,277,438	388,858	-----	$H_a: \hat{Y}_{ij} = a_i + bX_{ij}$
Regression (a ₁ ...a ₆ , b ₁ ...b ₆)	5	2,391,979	478,396	1.52	$H_0: \hat{Y}_{ij} = a_i + bX_{ij}$
Error 3	6	1,885,459	314,243	-----	$H_a: \hat{Y}_{ij} = a_i + b_iX_{ij}$

^aX = trunk cross-sectional area, Y = estimated total of nut clusters.

^bIndicates significance at 1 percent level.

Table 3.—Analysis of variance on the regression equations ^a

Source of variation	Degrees of freedom	Sums of squares	Mean Square	F-test	Hypothesis
Between groups	5	383,028	76,605	7.01	$H_0: \bar{Y}_i - \bar{Y}_k = 0$
Within groups	30	327,732	10,924	-----	$H_a: \bar{Y}_i - \bar{Y}_k \neq 0$
Total corrected sums of squares	35	710,760	20,307	-----	
Regression (a, b)	1	63,747	63,747	-----	$H_0: \hat{Y} = \bar{Y}$
Error 1	34	647,013	19,030	-----	$H_a: \hat{Y} = a + bX$
Regression (a ₁ ...a ₆ , b)	5	431,619	86,324	^b 11.62	$H_0: \hat{Y}_i = a + bX$
Error 2	29	215,394	7,427	-----	$H_a: \hat{Y}_i = a_i + bX$
Regression (a ₁ ...a ₆ , b ₁ ...b ₆)	5	30,794	6,159	.80	$H_0: \hat{Y}_i = a_i + bX$
Error 3	24	184,000	7,692	-----	$H_a: \hat{Y}_i = a_i + b_iX$

^aX = cross-sectional area of the primary scaffold, Y = estimated total nut clusters on the primary scaffold within trees.

^bIndicates significance at 1 percent level.

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

The correlation coefficient for the sum of CSA's for primary limbs with estimated total nut clusters was highly significant ($r = 0.95$). However, the correlation for trunk CSA with total nut clusters was not significantly different from zero at the 0.05 level.

Further study was done on the cost of obtaining these measurements in terms of time required for (1) walking from one tree to another, and (2) making 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). The time required at the tree for obtaining (1) the sum of the primary CSA's was about 3 minutes, and (2) the trunk CSA was about 1 minute. Thus, the time required for

both measurements was 4 minutes per tree for one person. These measurements need not be redone each year and could be used for about 4 years.

Tree Estimating Models—Regression and Ratio

To determine whether a ratio rather than a regression estimator should be used, one must satisfy the double requirements that (1) the correlations must be significant; i.e., r generally greater than $1/2 (S_x/\bar{X})/(S_y/\bar{Y})$ and

(2) the ratio of $\frac{b^2}{a^2} \frac{(1-f)}{n}$ must be greater than $\frac{\bar{X}^2 V(1/\bar{X})}{S_x^2}$

(4) (a and b are the parameters of the regression equation and $V(1/\bar{X})$ is the variance of the harmonic mean).

If the correlation is large enough, then the second criterion must be met. It is less binding since the

Y-intercept a can frequently be reduced by a simple transformation. For example, if the correlation is high and the slope b is large, but the intercept is also large (figure 6), a simple transformation of the X-variable can reduce the Y-intercept to zero (figure 7).

The regression estimator is not restricted by the value of a single Y intercept if a within-block model is used. A single translation of the regression as shown in figures 6 and 7 would be impossible if a new intercept were required for each block.

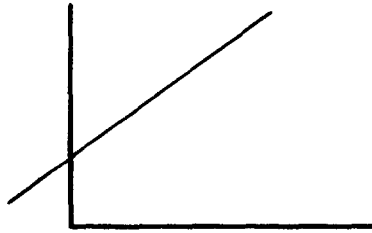


Figure 6

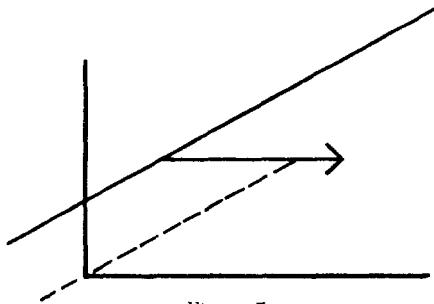


Figure 7

Tree Estimating Model Using Primary Limbs

The size of the primary limbs could range from about 2.0 up to about 20.0 square inches, insuring a wide range of sizes in the sample.

To evaluate which estimation procedure is more efficient for primary limbs, r was compared with $1/2 (S_x/\bar{X})/(S_y/\bar{Y})$. To compute the correlation coefficient, it was necessary to determine how the data should be combined. Table 3 shows the tests for combining within-tree data. The model testing procedures for this analysis were the same as those described previously for trees.

The model for utilizing primary limb data from the different blocks was found to be $\hat{Y}_{ij} = a_i + bX_{ij}$, where b is the average within-block slope for all data and a different intercept a_i for each i th block must be

computed. \hat{Y}_{ij} is total clusters on j th primary limb in i th block and X_{ij} is size of the j th primary in the i th block.

The correlation coefficient (assuming one average slope) can be computed from the values in table 4 by dividing the regression sum of squares by the within-group sum of squares and taking the square root of the quotient. The correlation coefficient is slightly larger than $1/2 (S_x / \bar{X}) / (S_y / \bar{Y})$. This relationship is based on an approximation of the mean square error of the classical ratio estimate. The inequality is approximate and if the correlation is high and the slope large (as for this case) the size information may still be helpful. The correlation computed was between X_{ij} (the size of the j th primary in the i th block) and the \hat{Y}_{ij} (estimated number of clusters on the same limb).

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

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

or,

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

Inserting the computed values for the variables, we conclude that a must be less than 64. The intercepts computed for the six blocks were: $a_1 = 42$, $a_2 = 125$, $a_3 = -7$, $a_4 = 111$, $a_5 = 327$, $a_6 = 12$. In three blocks (a_2 , a_4 , and a_5) the ratio estimator would have been more efficient and in the other three (a_1 , a_3 and a_6), the ratio estimator would be less efficient than the simple direct expansion. Furthermore, the intercept could not be changed by a single linear transformation because the intercepts varied so widely (-7 to 327). For this reason, a within-block regression estimator is better using the following model:

$$\hat{Y}_i = \bar{Y}_i + b(\bar{X}_{i1} - \bar{X}_{is}) = \bar{Y}_i + b\bar{X}_{i1} - b\bar{X}_{is}$$

where $\bar{Y}_i - b\bar{X}_{is}$ is the block intercept a_i , \hat{Y}_i is the double sampling estimate of the number of total clusters per tree in the i th block, \bar{Y}_i is the average of the direct expansion estimates for trees in the i th block, b is the overall regression coefficient, \bar{X}_{i1} is the average primary size for the block, and \bar{X}_{is} is the average size of the primaries sampled.

Since the coefficient of determination is 0.34 and the slope is significant, use of the primary limb size data should reduce the primary variance component by about

Table 4.—Within-block sums of squares used to compute correlation coefficient

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

$$r^2 = \frac{112,338}{327,732} = 0.343 \quad r = \sqrt{0.343} = 0.586 \quad (S_x / \bar{X}) = 0.608$$

$$(S_y / \bar{Y}) = 0.521 \text{ and } 1/2 (S_x / \bar{X}) / (S_y / \bar{Y}) = 0.583$$

one-third. Data on more blocks would help evaluate the reduction achieved by using CSA's of primary limbs in the estimation process.

Terminal Limbs Within Primary Limbs

The primary sample units (SU's) were subdivided into terminal SU's. This unit was defined as any limb with a CSA between 0.8 and 2.5 square inches. The average terminal SU had 50 nut clusters and took approximately 13 minutes to count. Two estimation schemes were studied: Equal probability selection with expansion by reciprocal of probability, and expansion using terminal size as an auxiliary variable in a ratio or regression estimate.

To determine which method of estimation was more efficient the same test discussed earlier was used. The *F*-value (15.46) for the second test was significant. The model using one average slope with a different intercept is also the best model for combining data for the terminal limb sample stage. For this analysis, $r = 0.20$, which is less than $1/2 (S_x/x) / (S_y/y) = 0.34$. Thus, the first criterion necessary for size to be used in the estimation procedure using a ratio estimator is not met. Neither the ratio nor the regression estimation scheme, which uses the terminal size, would reduce the variance because of the very low correlation. Therefore, if the terminal SU's are restricted in size from 0.8 to 2.5 square inches, then the simple unbiased estimator is more efficient than estimators using limb sizes in the estimation process.

Optimum Number of Trees, Primary Limbs, and Terminal Limbs

Two sample allocations were optimized: (1) optimum values for trees n , primaries within trees m , and terminals within primaries t , and (2) the optimum ratio of trees measured to trees counted.

Both optimizations assumed that all selections would be with equal probability with variance components estimated from sample data. The estimating model for the average tree within the k th block is:

$$Y_k = \frac{1}{n} \sum_{i=1}^n \frac{h_{ki}}{m} \sum_{j=1}^M \frac{M_{kij}}{t} \sum_{w=1}^t X_{kijw}$$

where

X_{kijw} = number of filbert clusters for the w th limb of the j th primary on the i th tree in the k th block

t = number of terminal sample units selected

M_{kij} = number of terminals on the j th primary in the i th tree in the k th block

m = number of primary sample units selected

h_{ki} = number of primary sample units on the i th tree in the k th block

n = number of sample trees per block.

Its associated variance formula is:

Total variance =

$$\frac{S_B^2}{k} + \frac{S_T^2}{kn} + \left(\frac{\bar{M}-m}{\bar{M}} \right) \left(\frac{S_P^2}{kmn} \right) + \left(\frac{\bar{T}-t}{\bar{T}} \right) \left(\frac{S_{Ter}^2}{knmt} \right)$$

and the appropriate cost function is:

Total cost =

$$(k) C_B + (kn) C_T + (knm) C_P + (knmt) C_{Ter}$$

where

k = number of blocks in sample

S_B^2 = variance component between blocks

C_B = cost of going from block to block (or block to home)

S_T^2 = variance component between trees

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

S_p^2 = variance component between primaries

C_p = cost of selecting one primary and breaking it into terminal sample units

S_{Ter}^2 = variance component between terminals within primaries

C_{Ter} = cost of selecting and counting one terminal

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

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

According to Snedecor and Cochran (5), the optimum values for t , m , and n are:

$$t = \sqrt{\frac{C_p S_{Ter}^2}{C_{Ter} S_p^2}}; \quad m = \sqrt{\frac{C_T S_p^2}{C_p S_T^2}}; \quad n = \sqrt{\frac{C_B S_T^2}{C_T S_B^2}}$$

The numerical values which were substituted are found in table 5. The optimum values rounded to integers are $n = 3$, $m = 1$, and $t = 2$. The next step is to find the optimum ratio of trees measured to trees counted. To optimize the ratio n'/n , again variance and cost functions are necessary. For this, a within-block function is needed and is as follows:

$$\text{Within-block variance} = \frac{S_T^2}{n} + \frac{S_p^2}{nm} + \frac{S_{Ter}^2}{nmt}$$

This must be changed to include double sampling at the tree level as follows:

Within-block sampling variance =

$$\frac{S_T^2(r^2)}{n'} + \frac{S_T^2(1-r^2)}{n} + \frac{S_p^2}{nm} + \frac{S_{Ter}^2}{nmt}$$

Table 5.—Summary of costs and variance components for the four stages of sampling

Source	Cost in minutes	Variance component
Blocks	150	118,113 = S_B^2
Trees	18	115,519 = S_T^2
Primaries	9	^a 91,334 = S_p^2
Terminal sample units	16	^a 554,293 = S_{Ter}^2

^aAdjusted for average finite population correction factors.

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, 4 minutes per tree, but could be used for 4 years so that an average of 1 minute per year was used, n' is the number of trees selected at random to measure, and r^2 is the coefficient of determination between the estimated quantity (total nut clusters) and the auxiliary variable (measure of tree size). In this study, we have recommended the sum of primaries as the covariate and assumed $r^2 = 0.7$; somewhat below the 0.95 observed in the section on the block estimating model.

The optimum ratio is found by forming the product of the variance and cost functions, differentiating with respect to n' and n , solving for each and forming the ratio.

The ratio before substitution is:

$$\frac{n'}{n} = \sqrt{\frac{S_T^2(r^2)(C_T + m C_p + mt C_{Ter})}{C_{T'} \left(S_T^2(1-r^2) + \frac{S_p^2}{m} + \frac{S_{Ter}^2}{mt} \right)}}$$

Using the variance components from table 5, and the cost values from above, n , m , and t as given indicate a ratio for n'/n of 3.2. Since three trees per block should be selected for counts, 10.2 is the optimum real number of trees which should be measured for the double sample. Operationally, 12 trees should be selected for measurements because 12 is a multiple of three and a rotation system for selecting the subsample for detailed counts and measurements could be worked out.

Selecting Trees, Primary Limbs, and Terminal Limbs

A sample of 12 trees should be selected at random for each block for obtaining the sum of primary limb CSA measurements for all primaries on each tree. These sums should be arrayed and a subsample of three trees systematically selected for identifying terminal limbs and making detailed cluster counts. The regression estimation technique should be used to adjust the estimated number of clusters for the subsample of trees for differences in tree size compared to the large sample of 12 trees.

For each of the three subsample trees one primary limb should be selected using equal probabilities (see figure 2). The selected primary limbs should be subdivided into terminal limbs and two of these selected for

making counts of nut clusters. Counts should be made by stripping all clusters from the sample terminals (see following section). This eliminates the need for a quality check on counts made. The number of clusters for the sample trees can then be estimated using a regression model where the direct expansion estimate for clusters (i.e., number of terminal limbs times nut clusters counted times number of primary limbs for tree) is adjusted for differences in the size of the sample primary and the average size for all primaries on the tree.

Errors in Counting

Information from quality checks from current survey procedures shows that the number of clusters on sample limbs has been undercounted. Data on counting accuracy were obtained for each person counting by having the supervisor strip all clusters from a subsample of terminal limbs which had first been counted with the usual on-tree counting procedure.

When the number of nut clusters missed (strip counts minus on tree cluster counts) is plotted against strip counts, the graphs indicate that a proportional relationship exists. The fitted line has a positive slope and goes approximately through the origin. This indicates that a factor could be applied to a limb count to adjust for undercounting. However, since the optimum terminal SU size is very small (CSA between 0.8 and 2.5 square inches), or generally between 1.7 and 2.0 percent of the tree, clusters for the entire limb can be stripped and counted. This eliminates some quality check work as most undercounts are usually associated with overlooking clusters partially hidden by leaves. Stripping eliminates most of this problem since counts are not made until the entire limb has been stripped and rechecked to see that no clusters were overlooked. Checks made to evaluate the accuracy of stripping found about 3 percent of the clusters not stripped. This compares with an undercounting of about 8 percent for prior methods.

Operational Survey

The first operational survey using the sampling techniques developed was completed in 1971. This survey required a sample of about 150 blocks, compared with 350 required when using the previous technique of

selecting a primary and counting all the nut clusters on it (about one-seventh of a tree). The derived sampling error from the new sample procedure was about 5 percent, 1 percent below the previous level, even though the number of blocks visited was reduced by 200. The new sampling procedure reduced survey costs by about 25 percent.

Summary and Conclusions

The sum of primary CSA's (1) is highly correlated with the estimated number of nut clusters per tree, (2) is inexpensive to obtain, and (3) can be used efficiently in a double sampling survey design. Primary limbs should be selected with equal probability and their size (CSA) used in the estimation process. Terminal limbs with CSA's between 0.8 and 2.5 square inches, selected with equal probabilities, should be used as sample units. The optimum sample allocation within a block is three trees, one primary limb per tree, and two terminal limbs per primary.

All nut clusters on selected terminals should be counted (stripped), picked and bagged. An independent quality count survey should be made a few days after the regular survey period to determine whether the proper limb was stripped and any nut clusters were missed. Bare tree photography for sample trees should be used for selecting primary and terminal limbs, and for the quality check survey.

References

- (1) Allen, Richard D., and Donald H. Von Steen. Use of photography and other objective yield procedures for citrus fruit. U.S. Dept. Agr., Statis. Rptg. Serv., June 1969.
- (2) Hansen, Morris H., William N. Hurwitz, and W. G. Madow. Sample survey methods and theory, vol. I. John Wiley and Sons, Inc., 1953.
- (3) Parr, Robert D., and Lyle D. Calvin. Research on objective forecasts of filbert production. Agr. Econ. Res., vol. 8, No. 3, pp. 92-95, July 1956.
- (4) Raj, Des. Sampling theory. McGraw-Hill, Inc., 1968.
- (5) Snedecor, George W., and William G. Cochran. Statistical methods. Sixth ed., Iowa State Univ. Press, 1967.