SAMPLING FOR OBJECTIVE ESTIMATES OF APPLE YIELDS

VIRGINIA, 1969

Research and Development Branch
Research Division
Statistical Reporting Service
U.S. Department of Agriculture
Washington D. C.

February 1973

SAMPLING FOR OBJECTIVE ESTIMATES OF APPLE YIELDS

VIRGINIA, 1969

By

Fred B. Warren

Research and Development Branch
Research Division
Statistical Reporting Service
U.S. Department of Agriculture
Washington D. C.

February 1973

CONTENTS

SUMMARY

A pilot survey in 30 different orchards in the Clarke County-Frederick County area of Virginia was conducted in 1969 to determine if selected objective procedures could be used operationally, and, if so, if their use would be economically practicable. Principal results from this survey were:

1. The correlation between counts of apples from photographs and the estimated number of apples on the same trees was quite high. However, the cost of obtaining and interpreting the photography was so high that, under present conditions, there is no economic advantage to using photography for this purpose.

2. Sample limbs can be selected from stereo photographs and identified in the field from photographs of the bare trees. Fifty-nine percent of the selected limbs were in the desired size range of 0.8 to 2.5 square inches cross sectional area (CSA). About 30 percent of the selected sample limbs had CSA's larger than 2.5 square inches. About 3 percent were larger than 7.5 square inches.

3. The size (CSA) of the primary limbs can be used advantageously to reduce sampling error or survey costs or both in a double sampling model where the CSA measurements are taken of the primary branches of a "large" sample of trees per orchard and counts of apples on selected limbs are taken from a "small" subsample of these trees.

This report also includes discussions of survey organization, of double sampling within trees to reduce sampling errors caused by barren limbs or by sample limbs of grossly unequal sizes, of non-geographic stratification in reducing sampling errors, and of estimating the proportion of apples counted in June which remain on the tree until harvest.

INTRODUCTION

Previous studies in one orchard in Virginia 1/ resulted in:

1. Development of a procedure for selecting sample limbs from stereo photographs of the bare trees.

2. Discovery of a significantly high correlation between the number of apples on a tree and

   (a) the number of apples which can be counted from projected color transparencies of the tree, and

   (b) the sum of the cross-sectional areas (CSA's) of the primary limbs of the tree.

1/ Warren, F. B. and Wigton, W. H., "Sampling for Objective Yields of Apples and Peaches, Virginia, 1967 and 1968," Statistical Reporting Service, U. S. Department of Agriculture.

These developments indicated that it should be possible to:

1. Implement a single-stage scheme of selecting terminal limbs in the office where, presumably, it would be better controlled.

2. Reduce the sampling variance or cost or both in a double-sampling sense by counting fruit on photographs of a comparatively large number of trees and correlating these counts with expanded counts of fruit on sample terminals from a subsample of the same trees, or by using the CSA's of primary limbs of a large number of trees in the same manner.

The objectives then were to determine from costs, variance components and correlations to be computed from a pilot survey in a major apple-producing region, whether these procedures can be used operationally and if they offer economic advantages over previously adopted procedures.

SAMPLE SELECTION

Blocks of apple trees used for this survey were selected from a list of trees reported on a census of growers in Clarke and Frederick Counties, Virginia. Blocks were selected with probabilities proportionate to the number of trees reported. Plantings reported to have been made within the last five years were considered to be not of bearing age and were excluded from the sample.

The list was arranged by age of planting within varieties. Thirty blocks were drawn systematically from this array. The distribution of these sample blocks by age and by variety of tree is given in Table 1.

Sample trees in the thirty blocks were selected, marked and photographed in April 1969, before the trees had started to leaf out.

A letter explaining the purpose and general nature of the project was sent to the owners of the selected blocks by the State statistician's office in Richmond, Virginia. A member of the Research and Development staff contacted each owner (or operator) to get permission to work in the selected planting, and to determine its location. In several cases, the orchard had been leased and the owner indicated the current operator. The block locations were recorded both on one-half inch to the mile county road maps and on 7-1/2 minute geological survey maps. Written instructions for getting to the planting and a large scale sketch which showed the approximate size and shape of the sample block, the location of any pollinators in the sample block and the nature of any adjacent plantings were also prepared. If the selected planting contained more than one thousand trees, the planting was divided into sections of from five hundred to one thousand trees and one of these was randomly selected for sampling.

Varieties (such as Golden Delicious) which are often used as pollinators for other varieties presented particular problems, both because they are usually

Table 1.--Distribution of sample apple blocks by variety and by age of planting, Virginia, 1969

Variety	Age of planting (years)					
	5-7	8-12	13-17	18-27	28+	Total
Grimes Golden		1				1
Jonathan					1	1
N.W. Greening					1	1
Red Delicious	1	2	1	1	1	6
Golden Delicious	1	1	1	1		4
Rome Beauty			1	1		2
Stayman			1	2		3
Winesap		1				1
York	2	1	1	3	4	11
Total	4	6	5	8	7	30

found interspersed with another varieties and also because the growers had not always reported them separately from the other variety on the 1967 fruit tree census. Consequently, the person making the initial contact, not only had to determine the location of any pollinators but also to find if the grower had reported the pollinators on the tree census. If another variety had been selected and the pollinators were reported separately, they were excluded from the tree selection process in that block. However, if the grower had reported the total number of trees in a planting as all of the principal variety, then all trees of all varieties in the planting were included in the tree selection.

After permission and the block location had been obtained, a two-man team visited the sample blocks to prepare more detailed maps of the sample block, to select sample trees, and to photograph a subsample of the trees.

The tree sampling procedure used was a three-stage nested design for tree selection and a modified one-stage design for selecting sample units within trees. All stages of sampling were with equal probabilities.

The first stage was a systematic sample of rows within the orchard and of trees within rows. The objective was to select a first stage sample of 32 to 40 trees which would be widely distributed over the orchard. The numbers of rows and trees per row were based on the shape of the individual block. Generally from four to eight rows per block and ten to four trees per row were sampled. Cross-sectional area (CSA) measurements were taken of all primary limbs on each first stage sample tree.

The second stage sample was a systematic subsample of nine trees from the list of 32 to 40 first stage sample trees. This subsample was taken after the trees on the list had been arrayed by the relative size of the sum of the primary CSA's. The ranking by size of CSA ensured that a wide range of tree sizes was selected, so it would be possible to determine if size of tree was important in the estimation procedure. A third stage sample of three trees was selected systematically from the second stage sample.

The trees in the third stage sample were photographed from opposite sides with a stereo camera and Kodachrome II film. The "sides" which were photographed were those which gave the best views of the branches of the tree.

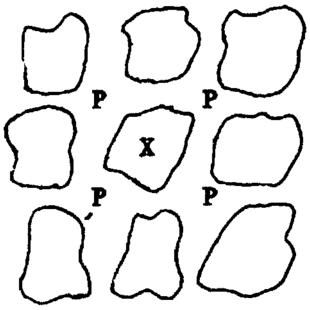
The stereo slides and black and white negative prints about 30 inches square made from the slides were used in the office to divide the tree into sample units. Sample units were defined as terminal limbs having CSA's of about 1.0 to 2.5 square inches and associated non-terminal limb sections. A non-terminal (or path) limb section was defined as being a section of limbs above one fork which were too thick to be classified as a terminal and which divided into at least two terminals and/or other path sections at the next highest fork. Each path section was uniquely assigned to the first terminal limb above it. The terminal limbs were numbered consecutively from one through t as they were identified. If two or more terminals emerged at the same location, path section(s) was assigned to the first terminal to be numbered.

Two clusters of three consecutively numbered sample units were selected systematically from the circular array of sample units. With the circular array, the sample unit having the largest number would be considered consecutively lower than sample unit number 1. For example, if a tree had a total of 26 sample units, there would be 26 possible clusters, including (25, 26, 1) and (26, 1, 2).

DATA COLLECTION

Field Procedures

During the last two weeks of June, all nine trees selected in the second stage samples were photographed from one "side" only. Due to the close spacing of apple trees in rows, normally there are only four positions from which the tree can be photographed (See Figure 1). The camera position selected for this photography was the one which was most nearly between the sun and the sample tree. The portion of the tree visible from this camera position was divided into approximately equal quadrants by a vertical pole and crossbar (See Figure 2). Each quadrant was photographed in color, using Kodachrome II film and a 35mm camera. Each photograph was identified by a clipboard fastened to the vertical pole which displayed the block number and the location of the tree in the block.



Legend

X is sample tree

P indicates possible camera positions

Figure 1.--Location of camera positions with respect to both the sample tree and to adjacent trees

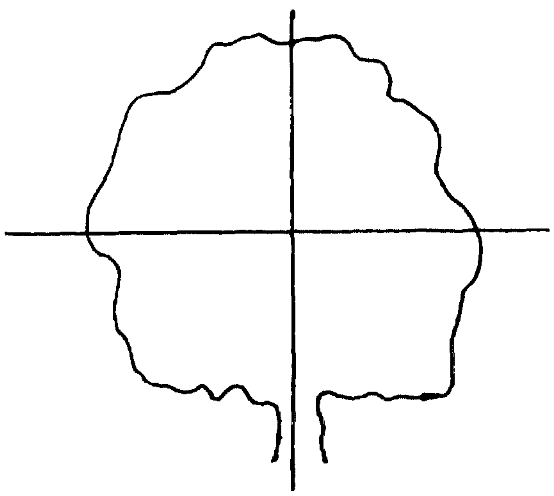


Figure 2.--Division of tree into quadrants by vertical standard and crossbar

The same day that the trees were photographed, the sample units selected from each third stage subsample tree were examined to see if they had any apples. This information was used to estimate the proportion (p) of sample units on the tree which had apples. CSA measurements were also taken of the terminal limbs included in these clusters. All apples on the "consecutively" lowest numbered sample unit in each cluster were counted if that limb had any apples. (If the cluster included sample units 25, 26, and 1, sample unit 25 would be counted). If the consecutively lowest numbered sample unit had no apples, either on the terminal or on any assigned path sections, apples were counted on the next lowest consecutively numbered sample unit in the cluster which had any apples. If none of the three sample units in a cluster had any apples, the count for the cluster would be zero.

Additional apple counts were taken in September on two trees in each of the ten Red Delicious or Golden Delicious sample blocks. This was a recount of the two clusters observed in June plus observations on enough additional clusters to bring the portion of the tree sampled up to about one-seventh to reduce the sampling error for the tree.

Photo Interpretation

Sets of photographs were assigned in a random sequence to one of four interpreters. Each set was composed of photographs of diagonally opposite (upper left and lower right, or lower left and upper right) quadrants on one side of a tree. To estimate the degree of variability between interpreters, a portion of the photographs were recounted by another interpreter. The assignment of photographs for recounting followed a randomized incomplete block design.

The Kodachrome transparencies were projected to approximately a 27 inch by 36 inch image on a viewing screen. This screen had been divided into a grid of 3 inch squares.

The counts from each slide were recorded on individual recording forms with the tree identification data, starting and ending photo count times and the date of the count. The count recording section of the form was laid out in the same grid type pattern as the viewing screen. The interpreter first recorded the position of the vertical and horizontal poles (as seen on the viewing screen) on the record sheet. He then made a cell by cell record of the apples observed in the assigned quadrant. Any apples which appeared in more than one cell were counted only in the lower or in the right hand cell.

ANALYSIS

Sample Limb Selection

The first step in the analysis was to determine if sample limbs selected from stereo photographs of the bare trees were of the desired size (0.8 to 2.5 square inches, CSA). A total of 540 terminal limbs (6 limbs per tree) were measured during the survey. Of this number, only 318 or 58.9 percent were within the desired range of CSA's (Table 2). Of the 222 remaining limbs, 59 were smaller than 0.8 square inches, CSA and 163 were larger than 2.5 square inches, CSA. The overall range in CSA was from 0.2 to 11.6 square inches.

Table 2.--Distribution of sample limbs by cross-sectional area, Virginia, 1969

Item	Cross-sectional area in square inches								Total
	Less than 0.8	0.8 to 1.5	1.5 to 2.5	2.6 to 3.7	3.8 to 5.0	5.1 to 7.5	7.6 to 10.0	10.1 to 12.5	
Number of limbs	59	172	146	92	36	20	13	2	540
Percent of total	10.9	31.9	27.0	17.0	6.7	3.7	2.4	0.4	100.0

Further analysis of the limb size data revealed a definite tendency for sample limbs to be larger in the older trees than in the younger trees (Table 3). Also, the average CSA of all sample limbs on trees which were at least 13 years old was larger than the desired maximum.

Table 3.--Distribution of sample limb size (CSA) by age of trees, Virginia, 1969

Age of tree	Limbs	Mean CSA	Standard error of mean	F value 1/
Years:	Number	Square inches	Square inches	
5-7	72	1.61	.21	*1.47
8-12	108	1.92	.14	**3.96
13-17	108	2.51	.26	0.26
18-27	144	2.67	.19	0.34
28 and older	108	2.85	.23	

1/ The computed "F values" test the hypothesis that any apparent difference between the average size of terminals in an age group and the next older age group is due to chance.

* and ** indicate that hypothesis was rejected at the 5 and 1 percent levels of probability, respectively.

Table 4.--Number of terminal size limbs per tree, by age groups, Virginia, 1969

Age of tree	Number of terminal size limbs per tree	
	Mean	Standard deviation
Years:		
5-7	17.75	8.83
8-12	31.50	9.94
13-17	46.80	18.31
18-27	56.04	26.72
28 and older	51.43	18.00
All ages	43.41	22.72

Three possible reasons for the failure to define terminal limbs of the desired size relate to the overall size of the older trees. These are:

(1) The apparent size of any object decreases as the distance to it increases. Photographs of the older (and larger) trees were generally taken at a greater distance from the tree. Also, the vertical and horizontal spread of the tree would be larger. Both factors would work to decrease the apparent size of branches, particularly at the top and sides of the tree. No attempt was made to correct for the increased distance from the tree in the sample unit definition.

(2) Many secondary limbs carry a large number of small branches, but have no well defined forks where both branches are large enough to be classified as terminals.

(3) On trees with a large number of terminals, the photo-interpreter might easily be tempted to not follow a branch as far as he could. There was a high positive correlation between the number of terminals identified on a tree and the average size of the sample terminals.

If the number of fruit per limb is correlated with the size of the limb, and the limb sizes are not approximately equal, then selection of limbs with equal probabilities will be less efficient than selection with probabilities proportionate to size (pps)--unless the size of the terminals is used as a covariate in the estimation process.

A system of sequential tests to determine if using the terminal CSA's as covariates in the estimation model was feasible indicated that by far the highest correlation between the number of apples per limb and the CSA was obtained with a regression model which used different intercepts and regression coefficients for each block (Table 5). This reflects some varying productivity levels of the different blocks either due to age, variety or management. However, a very large portion of the total sum of squares is explained by a common regression for all blocks.

Table 5.--Analysis of regression coefficients: CSA of sample limbs vs. apples per sample unit, Virginia, 1969

Source of variation	Degrees of freedom	Mean squares	F ratios	r values
Linear regression, $Y = a + bX$	1	116,556.04	**50.06	0.50
Mean, $Y = \bar{Y}$	153	2,328.18		
Linear regression with different intercepts for each block, $Y = a_i + bX$	28	4,900.07	** 2.80	0.59
Linear regression with a common intercept for all blocks, $Y = a + bX$	125	1,752.07		
Linear regression with different regression coefficients for each block, $Y = a_i + b_i X$	28	3,434.10	** 2.71	0.80
Linear regression with a common regression coefficient, $Y = a_i + bX$	97	1,266.54		

**In repeated sampling with these degrees of freedom an F value this large would occur by chance less than 1 percent of the time.

Direct Expansion Estimates

The limb sampling procedure used in this study selected two clusters of three consecutively numbered sample units on each tree. The presence or absence of apples was recorded for each of the three sample units. The number of apples was recorded only for the lowest numbered sample unit which had apples. This was done to determine if such a procedure would effectively reduce the within tree component of variance by reducing the number of zero fruit counts.

The direct expansion model for estimating the number of apples per tree from counts made on the first sample unit in each cluster would be

(1) Model 1: $\hat{X}_{1i} = \frac{N_i}{m} \sum_{j}^{m} X_{ij}$, $j=1,2,\ldots,m$

where N_i is the total number of terminals on the i^{th} tree,

m is the number of clusters sampled, and

X_{ij} is the count of apples on the first sample unit in the j^{th} cluster.

However, if the count of apples is taken from the first non-barren sample unit in the cluster, the proportion of sample units which are not barren must be observed. A separate proportion can be observed for each cluster, or a single proportion could be estimated for the entire tree. This gives the alternate formulas

(2) Model 2: $\hat{X}_{2i} = \frac{N_i}{m} \sum_{j}^{m} P_{ij} X_{ij}'$,

where the proportion of sample units with fruit (P_{ij}) is observed for each cluster, and

(3) Model 3: $\hat{X}_{3i} = \frac{N_i \hat{P}_i}{m'} \sum_{j}^{m} X_{ij}'$,

where the proportion of sample units with fruit is estimated for the entire tree

$(\hat{P}_i = \sum_{j}^{m} P_{ij}/m)$.

Model 2 may also be described as a two strata estimator within a cluster. One stratum would be assigned all limbs with no fruit. The second stratum would be assigned all count units having at least one fruit.

$$\hat{X}_{2i} = N_i \frac{1}{3m} \sum_{j}^{m} (m_{j1}(0) + m_{j2} X_{ij}), \text{ or}$$

$$\hat{X}_{2i} = N_i \frac{1}{m} \sum_{j}^{m} \frac{m_{j2}}{3} X_{ij} = \frac{N_i}{m} \sum_{j}^{m} P_{ij} X_{ij},$$

where m_{j1} and m_{j2} are the number of limbs in the two strata for the j^{th} cluster and $m_{j1} + m_{j2} = 3$.

In Models 2 and 3, X_{ij}' is the number of apples counted on the first non-barren sample unit in the cluster if there was at least one sample unit with fruit. Otherwise, $X_{ij}' = 0$. Also in Model 3, m' is the number of clusters where $X_{ij}' > 0$.

Model 3 involves the product of two random variables, the estimated proportion of non-barren sample units per tree and the sum of the fruit counted in the first non-barren (if any) sample unit in each of the m clusters. As such, it would have a lower within tree variance than Model 1 only if the correlation between these two variables was negative.

The relative efficiency of these models was computed as the ratio of the variances which would be expected under an optimum sample allocation (Table 6). For the given costs, Model 2 was not quite as efficient as Model 1. The lower within tree component of variance for Model 2 was more than canceled by the increased cost of obtaining the supplementary data.

The expected variance of Model 3 was 12 percent higher than Model 1. Unlike direct expansion Models 1 and 2, Model 3 includes a ratio estimate, the proportion of non-barren limbs on the tree. Such a model would result in a reduced within tree variance only if the proportion of non-barren limbs was negatively correlated with the number of apples per non-barren limb. In this study, the observed correlation was +0.32.

Therefore, the data observed in this study indicates that:

(a) there would be no advantage in using Model 2, which is a more complex procedure, in preference to Model 1, and

(b) that Model 3 is definitely inferior to both Model 1 and Model 2.

The expected variance of an estimate of total fruit per tree from a survey is computed as

$$S^2 = \frac{s_b^2}{n_b} + \frac{s_t^2}{n_b n_t} + \frac{s_w^2}{n_b n_t n_w},$$

where the s_i^2 are estimated components of variance, and the n_i are the number of observations to be taken at each stage of subsampling. For Model I,

$$S^2 = \frac{1,732,000}{186.1} + \frac{739,000}{(186.1)(1.4)} + \frac{8,465,000}{(186.1)(1.4)(5.5)}$$

$$= 17,642, \text{ the result in Table 6 for the optimum allocation.}$$

Practically, the optimum number of samples at each stage of sampling must be rounded to a whole number. Therefore, the expected value of S^2 could be

$$S^2 = \frac{1,732,000}{186} + \frac{739,000}{(186)(1)} + \frac{8,465,000}{(186)(1)(6)} = 20,870$$

Table 6.--Variance components, expected costs, and optimum allocations of sample units under Models 1, 2, and 3, Virginia 1969

Item and level of sampling	Model 1	Model 2	Model 3
Estimated apples per tree	1,589	1,534	1,686
Variance components:			
Between blocks (000)	1,732	1,732	1,732
Between trees within blocks (000)	739	739	739
Within trees (000)	8,465	8,204	9,987
Cost per unit 1/:			
Blocks (dollars)	12.70	12.70	12.70
Trees (dollars)	2.50	2.50	2.50
Clusters (dollars)	1.30	1.45	1.50
Optimum allocation 2/:			
Blocks	186.1	183.7	172.2
Trees per block	1.4	1.5	1.5
Cluster per tree	5.5	5.1	5.7
Estimated variance per tree	17,642	18,137	19,847
Efficiency relative to Model 1 (Percent)	100	103	112

1/ The listed costs include one-third of the original cost of locating sample blocks and trees, photographing sample trees, and identifying terminals on the photographs.

2/ The optimum allocations were computed using a total survey cost of $5,000.

Double Sampling - CSA's of Terminal Limbs

The high correlation (Table 5) between the observed CSA's of the sampled terminal limbs and the number of fruit per sample unit indicates that CSA's might be used in a double-sampling model to reduce the sampling error of the estimated number of fruit. This would be true only if the coefficient of correlation (r) between the terminal CSA's and the number of fruit counted on those sample units is greater than the term

(4) $\sqrt{\frac{4C_1C_2}{C_1+C_2}}$ 2/ , where

2/ William G. Cochran, Sampling Techniques, John Wiley and Sons, Inc., New York, New York, Second Edition, 1963, p. 337.

C_1 is the average cost of counting fruit on a sample unit, and

C_2 is the average cost of obtaining a CSA measurement on a terminal limb.

When CSA measurements are obtained from three terminals per cluster, C_1 is \$1.28 and C_2 is \$0.15, so that the coefficient of correlation must be greater than .61 if double sampling for CSA's is to be effective.

This condition is met only for the case where a different slope and a different intercept is computed for each block (Table 5). Since different intercepts are required, all of the intercepts cannot be zero. Therefore, the double sampling ratio model would be biased and we considered the regression model

(5) Model 4: $\hat{X}_{4ij} = n_i (X_{ij} + b (\overline{Y}_i - Y_{ij}))$, where

n_i and X_{ij} are as defined for Model 1,

$\overline{Y}_i$ is the average CSA of all terminal limbs in all sample clusters on the tree,

Y_{ij} is the CSA of the terminal limb in the first sampling unit of the j^{th} cluster, and

b is the coefficient of regression computed from the fruit counts (X_{ij}) and the corresponding terminal CSA (Y_{ij}) measurements in each block.

A second double sampling model comparable to Model 2 (2), would be

(6) Model 5: $\hat{X}_{5ij} = n_i \hat{P}_i (X_{ij}' + b'(\overline{Y}_i - Y_{ij}'))$, where

the n_i, $\hat{P}_i$, X_{ij}', and $\overline{Y}_i$ have been defined previously. Y_{ij}' is the

CSA of the terminal in the first sample unit in

Variance components, costs, and optimum allocations of sampling units for these double-sampling models given in Table 7 indicate that Model 1 would be only 58 percent as efficient as Model 4 and 59 percent as efficient as Model 5.

However, the optimum allocation would result in observing only 3 or 4 sample units (clusters) in each block. Since a regression coefficient (b) must be computed for each block, a larger number of observations in each block may be desirable. This would result in some reduction in the efficiency of the double-sampling models. However, for moderate increases in the number of observations per block, the double-sampling models would still be more efficient than Model 1 (Table 8).

Table 7.--Variance components, expected costs, and optimum allocations of sample units under double-sampling, Virginia 1969

Item and level of sampling	Model	
	4	5
Variance components:		
Between blocks....................(000)	1,496	1,496
Between trees within blocks.......(000)	739	739
Within trees......................(000)	1,474	1,158
Cost per unit 1/:		
Blocks.......................(dollars)	12.70	12.70
Trees........................(dollars)	2.50	2.50
Clusters.....................(dollars)	1.43	1.58
Optimum allocation 2/:		
Blocks..................................	238.2	241.8
Trees per block.........................	1.6	1.6
Clusters per tree.......................	1.9	1.6
Expected variance of estimate............	10,282	9,999
Relative efficiency of estimate 3/....(%)	58	57

1/ Listed costs include one-third of the original costs of identifying sample blocks and trees, of photographing sample trees, and of identifying terminals on the photographs.
2/ Allocation was optimized using an allowable survey cost of $5,000.
3/ Relative to Model 1.

Table 8.--Sub-optimal allocation of sampling units under Model 4 with the constraint that the total number of observations per block be a specified number, Virginia 1969

Item	(Clusters per tree)x(Trees per block) =			
	6	8	9	12
Sub-optimal allocation:				
Blocks........................	190	171	151	133
Trees per block...............	2	2	3	3
Clusters per tree.............	3	4	3	4
Expected variance of estimate...	11,111	11,986	12,623	14,024
Relative efficiency....(Percent)	63	68	72	79

Double Sampling - Apple Counts From Photographs

The coefficient of correlation (r) between the number of apples counted from color transparencies of diagonally opposite quadrants of one side of a tree with the expanded limb counts (X_5) from the same trees was 0.68 when a single regression coefficient was used for all trees in all blocks (Table 9). A sequential analysis by age groups showed that using different intercepts for each age group did not improve the correlation materially, but that using a different regression coefficient for each age group did increase the correlation significantly, to 0.80. The individual regression coefficients varied from a maximum of 13.32 and to a low of 2.63 (Table 10). The difference in regression coefficients may be partially related to the relative density of the foliage for trees at different ages.

Even though the correlations between photo counts and the expanded limb counts are significantly different from zero, there is still the question "How large must the coefficient of correlation be to offset the increased cost of obtaining the additional information?" This quantity is computed as

$$r > \sqrt{\frac{4C_1C_2}{(C_1+C_2)}}$$

where C_1 is the cost per tree of the photo counts, and
C_2 is the cost per tree of the expanded limb counts.

Assuming that the expanded limb counts are estimated using two clusters per tree, C_2 would be \$5.35. If photo counts were of two diagonally opposed quadrants on one side of the tree, C_1, the cost of the photography and of the interpretation would be \$3.53. These values indicate that the coefficient of correlation would need to be larger than 0.98 for a double sampling model using photo counts to be justified.

It might be possible to substantially reduce the cost of the photo counts in some way. If so, what is the maximum value that C_1 can take? This can be computed by expanding and rewriting the above formula as a quadratic in C_1

$$C_1^2 + (2-4/r^2)\, C_2C_1 + C_2^2 = 0$$

Substituting for C_2 and r, we have

$$C_1^2 - 22.7375\, C_1 + 28.6225 = 0$$

Solving the quadratic, C_1 = \$1.34. This is 2 cents less than the estimated cost of selecting sample trees, taking photographs and processing the film of \$1.36. This indicates that the use of photographs in estimating apples will be impracticable under the stated cost conditions.

Table 9.--Analysis of regression coefficients: Sum of photo counts (X) vs. expanded limb counts (Y), and coefficients of correlation, Virginia 1969

Source of variation	Degrees of freedom	Mean squares	F ratios	r values
Linear regression, $Y = a + bX$	1	144,584,571	**72.3	.68
Mean, $Y = \overline{Y}$	86	2,000,452		
Linear regression with different intercepts for each age group, $Y = a_i + bX$	4	2,705,368	1.38	.70
Linear regression with a common intercept for all ages, $Y = a + bX$	82	1,966,065		
Linear regression with different regression coefficients for each age group, $Y = a_i + b_iX$	4	11,775,666	**8.05	.80
Linear regression with a common regression coefficient for all ages, $Y = a_i + bX$	78	1,463,009		

** In repeated sampling, an F value this large would occur by chance less than one percent of the time.

Table 10.--Regression coefficients and coefficients of correlation between photo counts and expanded limb counts by age groups, Virginia 1969

Age of trees	Sample size (n)	Coefficient of correlation (r)	Regression coefficient (b)
5 to 7 years	12	0.86	7.49
8 to 12	17	0.63	5.49
13 to 17	15	0.89	13.32
18 to 27	23	0.69	6.56
28 and older	21	0.59	2.63

Double Sampling - CSA's of Primary Limbs

Correlation between the sum of the primary CSA's and the estimated number of apples per tree were computed for all trees (Table 11), for trees within age groups (Table 11), and for trees within each block (Table 12). The correlation over all trees was significant at the 5 percent level ($r = .33$, 88 d.f.). A highly significant improvement in the joint correlation ($r = .67$) was obtained by using different intercepts and regression coefficients for each age group. The correlation within the oldest age groups (trees at least 28 years old) was not significantly large (Table 13). Still greater efficiency ($r = .80$) resulted from using a model which had a different intercept but the same regression coefficient for each block.

The inequality $r = \sqrt{4C_1C_2}/(C_1+C_2)$ again was used to determine if the sum of the primary CSA's would be useful in a double sampling model. Here, C_2 is still \$5.35 but C_1 is only \$0.56, so the computed $r = 0.59$. The observed correlation is much higher than 0.59 so this variable would be suitable for a double sampling model.

The formula used to determine the optimum ratio of trees for which sample limbs are to be counted to those for which the primary limbs will be measured was

$$\frac{n_2}{n_1} = \sqrt{\frac{1-r^2}{r^2} \cdot \frac{C_1}{C_2}}\,, \text{ 3/}$$

$$= \sqrt{\frac{.20}{.80} \quad \frac{.56}{5.35}} = .16$$

Therefore, the primary CSA's should be measured on 6 trees for each tree on which sample limbs counts are made.

The double sample estimate of the average number of apples per tree ($\hat{X}'$) for a particular block then would be

$$\hat{X}' = \frac{1}{n_2} \sum_{1}^{n_2} \hat{X}_i + b\,(\bar{Y} - \bar{Y}'),$$

where n_2 is the number of trees where sample limbs were counted,

$\hat{X}_i$ is the direct expansion estimate of the number of apples per tree (possibly adjusted for terminal CSA or the estimated proportion of barren limbs),

b is the least squares estimate of the common regression coefficient computed over all blocks,

---

3/ Des Raj. Sampling Theory, McGraw-Hill Book Co., New York, 1968, p. 92.

Table 11.--Analysis of regression coefficients by age groups: Sum of primary CSA measurements (X) vs. estimated apples per tree (Y), and coefficients of correlation (r) Virginia 1969

Source of variation	Degrees of freedom	Mean squares	F ratios	r values
Linear regression, $Y = a + bX$.....	1	45,141,434	**11.12	.33
Mean, $Y_{ij} = \bar{Y}$....................	88	4,039,989		
Linear regression with different intercepts for each age group, $Y_{ij} = a_i + bX$..................	4	19,775,234	** 3.97	.56
Linear regression with a common intercept for all ages, $Y = a + bX$......................	84	3,311,644		
Linear regression with different regression coefficients for each age group, $Y = a_i + b_iX$..................	4	14,239,956	** 5.15	.67
Linear regression with a common regression coefficient for all ages, $Y = a_i + bX$...............	80	2,765,229		

** In repeated sampling, an F value this large would occur by chance less than one percent of the time.

$\bar{Y}$ is the average of the sum of the primary CSA's of all measured trees in that block,

$\bar{Y}'$ is the average of the sum of the primary CSA's of those trees where sample limbs were counted.

This procedure does not exclude barren trees ($\hat{X}_i = 0$) from the computations for b. To do so would require making a separate estimate of the proportion of barren trees in each block.

Table 12.--Analysis of regression coefficients by blocks: Sum of primary CSA measurements (X) vs. estimated apples per tree (Y), and coefficients of correlation (r), Virginia 1969

Source of variation	Degrees of freedom	Mean squares	F ratios	r values
Linear regression, $Y = a + bX$.....	1	45,141,434	**11.12	.33
Mean, $Y = \bar{Y}$....................	85	4,059,989		
Linear regression with different intercepts for each block, $Y = a_i + bX$..................	28	7,353,287	** 3.01	.80
Linear regression with a common intercept for all blocks, $Y = a + bX$.....................	57	2,331,250		
Linear regression with different regression coefficients for each block, $Y = a_i + b_iX$........	28	2,997,849	1.06	.92
Linear regression with a common regression coefficient for all blocks, $Y = a_i + bX$............	29	1,903,203		

** In repeated sampling, an F value this large would occur by chance less than one percent of the time.

Table 13.--Regression of estimated number of apples per tree on sum of primary CSA measurements, by age groups, Virginia 1969

Age of trees	Degrees of freedom	r values	b values
5 to 7 years.....................	11	.76	56.10
8 to 12..........................	17	.46	35.69
13 to 17.........................	14	.78	46.74
18 to 27.........................	23	.60	20.19
28 and older.....................	20	.06	0.94

Non-geographical Stratification

The Statistical Reporting Service routinely uses stratification in many of its surveys. The stratification is always according to some characteristic of the individual sample unit whether it be geographic, size of operation, or some other feature. The division of the sample (and by implication, the universe being sampled) into strata may be done before the survey on the basis of "known" attributes, or after the survey on the basis of information obtained during the survey. These alternate procedures are defined as pre- and post-stratification.

The purpose of stratifying a sample may be to obtain separate estimates for the different strata or to reduce the sampling error of the combined estimate or both. Stratification will always reduce the sampling error of the combined estimate whenever there is an appreciable difference in the average levels of the sample values in the different strata. The sampling error will be reduced more if the proportion of sample units in each stratum is known beforehand, rather than estimated from the survey. To demonstrate this fact, consider that the combined estimate of the mean, $\bar{\bar{X}}$, is computed as:

$$(1) \quad \bar{\bar{X}} = \sum_{h=1}^{m} P_h \bar{X}_h$$

if $P_h = N_h/N$. However, if P_h is estimated as the proportion of sample units in the stratum h from some sample survey, i.e., $P_h = n_h/n$, then P_h becomes a random variable with mean P_h and variance $S^2 p_h$. Accordingly, equation (1) would be rewritten as:

$$(2) \quad \bar{\bar{X}} = \sum_{h=1}^{m} \hat{P}_h \bar{X}_h$$

The importance of the P vs. $\hat{P}$ lies in the resulting variance formulas. The variance of the estimate in (1) is:

$$(3) \quad S_{\bar{\bar{X}}}^2 = \sum_{h=1}^{m} P_h^2 \, S_{\bar{x}_h}^2$$

and the variance in (2) is:

$$(4) \quad S_{\bar{\bar{X}}}^2 = \sum_{h=1}^{m} (P_h^2 \, S_{\bar{x}_h}^2 + \bar{X}_h^2 \, S_{p_h}^2)$$

$$= \sum_{h=1}^{m} P_h^2 \, S_{\bar{x}_h}^2 + \sum_{h=1}^{m} \bar{X}_h^2 \, S_{p_h}^2$$

Therefore, (4) will always be larger than (3) by the quantity $\sum_{h=1}^{m} \bar{X}_h^2 \, S_{p_h}^2$.

The sampling error can be reduced still further if

(a) there is significant difference between the between unit variances within the different strata, and

(b) the samples are allocated to the various strata by a system of optimum sample allocation, e.g., in proportion to the quantities $N_h S_h / \sum_{h=1}^{m} (N_h S_h)$, h = 1,2,...,m.

Optimum allocation of primary sample units to the different strata must be done before the survey, i.e., to a pre-stratified universe.

Characteristics to be examined for possible use as bases for stratification of the sample units should be readily identifiable. There should also be a logical reason for believing that a division of the universe by such characteristics would produce subsets of sample units which would have significantly different means or between unit variances or both for the value being estimated.

The age of the trees in the individual plantings could satisfy both requirements. Information on the age of plantings normally would be obtained as part of the tree census needed as a precursor to any sample survey of trees. Also, comparatively young and comparatively old trees would, on the average, be less productive than trees of an intermediate age. Furthermore, there might also be greater variations in productivity among the very young or very old trees than among those of medium age.

The use of varieties as an alternative basis for stratification is also included in this discussion even though it would not appear, a priori, to be as effective as age. Any variety that had proved to be significantly less productive would, unless it had other very strong characteristics in its favor, tend to be replaced by more productive varieties.

The estimated average number of apples per tree for each age group and the standard deviation of the block estimates of the average within each age group are listed in Table 14. (This particular choice of age groups for the stratum was dictated by the form in which the tree census was conducted. That is, growers reported trees as having been planted in one of several periods, e.g., 1962-64, 1957-61, etc.). Even though there is a wide variation in the estimated average number of apples among the different age groups, the between block variation within the age groups was so large that neither an F Test of overall differences between means, nor a more sensitive t-test of differences between means in adjacent age groups produced any results that were significant at the five percent level. However, Bartlett's test for homogeneity of variance between the age groups resulted in a chi square value (X^2 = 11.9, 4 df) which was significantly large at the five percent level.

These findings indicate that stratifying the sample by these age groups and allocating samples to the strata in proportion to the number of trees in each would result in at best only a minor reduction in the variance of the combined estimate. A more substantial reduction in the variance would occur if the samples were allocated optimally, i.e., in proportion to the product $N_h S_h$. To verify these assumption, the sample allocations and other data in Table 6 were used to construct simulated standard errors of estimated fruit per tree under these conditions (Table 16).

The simulated standard errors indicate that stratification by age groups with proportionate allocation of samples would effect only about a 13 percent reduction in the variance of the combined estimate. However, the same stratification with an optimum allocation of samples would reduce the variance by about 44 percent.

The stratification by variety groups examined in this report consisted of the three major winter varieties (York, Red Delicious, and Golden Delicious) in this area, all other winter varieties combined, and all fall and summer varieties combined (Table 15). As expected, there was not a large difference between variety groups in the estimated average number of apples per tree ($F = 0.2$; 4,25 df). Also, Bartlett's test for homogeneity of variance yielded a chi-square value of 6.56 with four degrees of freedom. A chi-square value larger than this would be obtained by chance at least 15 percent of the time in repeated sampling.

The data in Table 16 indicate that stratification by these variety groups and a proportionate allocation of samples to the stratum would result in a larger standard error for the estimated mean than would no stratification at all. Using an optimum rather than a proportionate allocation would reduce the standard error slightly. However, the indicated gains would be so small that such a procedure would hardly be worthwhile.

Table 14.--Estimated apples per tree, standard deviations, 1/ and sample allocations by age groups, Virginia 1969

Age groups	Apples per tree	Standard deviation (S_h)	Trees in group: Number (N_h)	Trees in group: Percent of total	Sample allocation: Proportionate	Sample allocation: Optimum
	Number	Number			Number	Number
Years:						
5-7	593	672	49,946	10.0	3	1
8-12	1,430	724	105,446	21.2	7	3
13-17	2,347	1,806	69,191	13.9	4	6
18-27	2,172	2,356	121,646	24.4	7	13
28 and older	593	992	152,200	30.5	9	7
Total			498,429	100.0	30	30

1/ The standard deviation for each age group is the square root of the variance of the between individual block estimates of the average number of apples per tree for that age group.

Table 15.--Estimated apples per tree, standard deviations 1/ and sample allocations by variety groups, Virginia 1969

Variety groups	Apples per tree	Standard deviation (S_h)	Trees in group: Number (N_h)	Trees in group: Percent of total	Sample allocation: Proportionate	Sample allocation: Optimum
	Number	Number			Number	Number
York	1,662	2,276	173,347	34.8	11	15
Red Delicious	1,338	1,046	103,138	20.7	6	4
Golden Delicious	1,873	1,934	70,282	14.1	4	5
Other winter	1,255	1,040	107,745	21.6	6	4
Fall and summer	960	985	43,867	8.8	3	2
Total			498,429	100.0	30	30

1/ The standard deviation for each variety group is the square root of the variance between individual block estimates of the average number of apples per tree for that variety group.

Proportion of Apples Remaining at Harvest

Apples on the same limbs of two trees in each of the ten Red and Golden Delicious sample blocks were counted in June and again in September. Analysis of these counts revealed that:

(1) The average loss of apples from June to September was 11.7 percent, i.e., the September/June) ratio was .883.

(2) The proportion of apples lost was linearly independent of the number counted in June. The linear coefficient of regression between apples counted in June and in September (.899) was not significantly different from the (September/June) ratio .883 ($F = 0.51$; 1,38 df). This would imply that the intercept of the regression line is also zero, so that there would be no need to use the regression model rather than the ratio model in predicting loss of apples.

(3) Variance components computed from the ratios of individual limb counts show that there is slightly more variation in rate of loss between individual branches within trees than between trees, and much more variation between trees within blocks than between blocks.

Table 16.—Simulated standard errors of the estimated number of apples per tree under different types of stratification and types of sample allocation, Virginia 1969

Type of stratification	Type of sample allocation	Standard error of: estimated mean	Relative efficiency 1/
			Percent
None	Systematic 2/	299	---
By age groups	Proportionate	279	113
	Optimum	248	144
By variety groups	Proportionate	307	94
	Optimum	291	104

1/ The relative efficiency of the stratified sampling plans compared with unstratified sampling is computed as:

$$100 \text{ X } \frac{(n_1 + 1)(n_2 + 3)s_2^2}{(n_2 + 1)(n_1 + 3)s_1^2}, \text{ where}$$

s_1^2 is the variance of the estimate from the stratified sample

n_1 is the degrees of freedom $(n - k)$ of s_1^2

s_2^2 is the variance of the estimate from the unstratified sample, and

n_2 is the degrees of freedom $(n - 1)$ of s_2^2

Reference: "Principles and Procedure of Statistics," Steele and Torrie, McGraw-Hill Book Co., Inc., 1960.

2/ Plantings were arranged by age group within varieties. The sample was taken systematically, with probabilities proportional to the size of the plantings, from this array.

Table 17.--Analysis of variance and variance components for ratios of apples in September to apples in June, Red Delicious and Golden Delicious apples, Virginia 1969

Source of variation	Degrees of freedom	Mean squares	F ratios	Variance component
Between blocks........	9	.03434	1.037	.00031
Between trees within blocks.......	10	.03312	2.484	.00989
Between limbs within trees........	20	.01333		.01333

RECOMMENDATIONS

Future work in developing objective estimation procedures for apples should include the following items:

(1) A determination of the feasibility of a two-stage limb selection procedure similar to that now used for filberts. This study should include cost effectiveness comparisons with both the Jessen multiple stage random path procedure and with the one stage equal probability procedure used in this project.

(2) A study of factors which would affect both the proportion of marketable apples left on the tree at harvest and the size (volume) of such apples. Some of these factors would vary by years so this should be set up as a multiple year project. This project probably could build on the sizing curves developed by Agricultural Research Service (ARS) of USDA.

APPENDIX

Virginia Apple Project - Sample Block Data Sheet

Owner: Name ____________________ Sample No. ____________________

Address ____________________ Date ____________________

Person to Contact: Name ____________________

Address ____________________

Telephone ____________________

Orchard Identification

1. Orchard No. ____________________
2. Acres in Orchard ____________________
3. Location ____________________

____

Block Identification

Variety Name ____________________

Age (Year) ____________________

Number of Trees in Age Group ____________________

	Number of Trees in Separate Blocks in Age Group		
		No.	Cum
Sample Tree in Age Group ________	1.	________	________
Block to be Sampled ________	2.	________	________
No. of Trees ________	3.	________	________
No. of Rows ________	4.	________	________
Trees per Row ________	5.	________	________
	6.	________	________

Comments

____

____

____

Fruit Counting Procedures
Virginia Apples - 1969

Objectives:

1. To estimate the proportion of sampling units on a tree which have any fruit.
2. To estimate - from counts of fruit on designated sampling units - the total number of fruit on the sample trees.
3. To evaluate the efficiency of the tree mapping procedures from the Itek prints and stereo slides.

Definitions:

Sampling Units: A terminal branch (or branches) and its associated paths (if any). Each sampling unit must be small enough to count without taking too much time.

Terminal Branch: Branch with a thickness as measured on the Itek prints as generally between 1/16" and 3/16" from which no other major branches emerge.

Path: A non-terminal section of a limb. Either terminal or non-terminal branches emerge from it.

Primary Limb: All limbs at first branching of trunk.

Cluster: A group of three sampling units that are numbered successively (i.e.)(3,4,5) etc. The first limb is the first selected terminal of the cluster.

Itek Print: A negative print made from the color slide that was used to divide the tree into sampling units.

Orchard Location:

You will be given a list of the sample orchards with directions for getting to them from major road intersections. The orchard locations are also located approximately on county highway maps and more precisely on geological survey 7-1/2" quadrangle maps. There is also a sketch of the selected block which is to a scale of 1 tree to a square (except for particular instances where pollinators have been planted more closely).

Sample Trees:

The location of the trees to be counted is given in terms of row and tree coordinates with respect to the starting corner. The coordinates of the trees are on the sample selection sheet for the block and on the count sheets for the individual trees. Nine trees in each sample block have been marked with red or yellow ribbon around the trunk of the tree. Only three of these trees are to be counted so you will need to count the trees and rows carefully to make certain that the counts are made on the proper tree.

The time that you arrive at each sample tree is to be recorded in the space provided in the upper left-hand corner of the count record form.

<u>Sample Units</u>:

1. For most trees, the location of the base of the terminal limbs and the location of the associated path sections are marked on the Itek prints for the tree. All terminals are numbered and all path sections are lettered. You will probably need to use the stereo viewer and slides of the tree to find the designated sample limbs. The orchard number, row and tree coordinates, and the direction from which the picture was taken is written in the left-hand margin of the Itek print.

   There will be two clusters of three sample units listed on the count form for each tree. The first unit listed for each cluster will be marked at the base of the terminal limb with orange ribbon. The path sections for that sampling unit will be marked with yellow ribbon.

2. Fill in the first section of form. This section is to find out how many sampling units in the two clusters of three limbs have fruit. As soon as the first fruit is found on either the terminal or its associated path (if it has any) - check yes. If no fruit is found on the terminal or its path, check no. This is to be done for all three sampling units in the two selected clusters. You will also measure the cross-sectional area of each terminal near its base and record the measurement to the right of the "no" column.

3. The fruit count section on the form is next. This section is to record accurate counts for a terminal limb. Counts will also be made for any path sections associated with the selected terminals. If the first sampling unit in the cluster had no fruit (i.e. marked no above) then the next limb is chosen for the counts. The path and the terminal in the second row of cluster are taped as before with orange and yellow tape and counts are made. In a few cases the third limb might be needed. The form has three columns, the first column identifies the terminal or path being counted; the second column numbers the subunit (spur) of that terminal (or path) and the third column gives number of fruit counted. Place a white tag at the base of each subunit. If there are more than 10 apples on a spur, place additional tags after the 10th, 20th, etc. apples. All counts will be verified by the other member of the counting team.

4. The third part of the form is for the purpose of evaluating the tree divisions. The path sections marked in this section must be checked

for terminal limbs missed in the mapping. If a primary limb is visible from more than one Itek print, both must be checked if a terminal limb is not shown on the one print. At this point, our definition of a terminal limb is a limb of greater size than the terminal to which it has been assigned.

June Photography Instructions

Objectives:

To photograph one side of nine selected apple trees in each block.

Field Procedures:

1. The nine trees to be photographed for each block are identified by block number, row number, and tree number in the row. They are also marked with red or red and yellow ribbon. Photographs will be taken from the side of the tree which (a) is specified on the form, (b) keeps the sun to your back as much as possible.

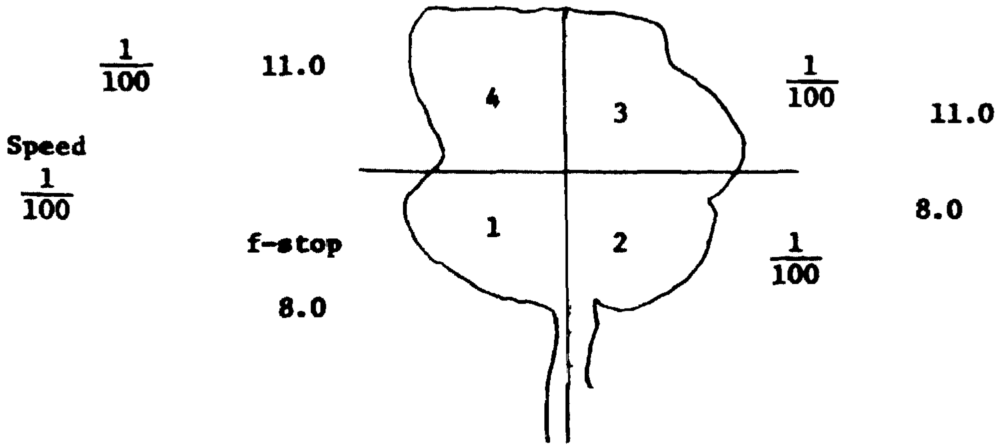
2. Once the side has been selected, place the aluminum poles in the form of a cross in front of the tree so as to divide the leaf surface of the tree into four nearly equal sections.

3. The light meter setting must be taken next to tree with camera pointed toward the base of the tree. This insures that fruit in the shadow of the leaves will be properly exposed. In order to assure enough depth of field, the f-stop should be placed at 8 or 11.

4. Set the camera on a tripod approximately 5 feet 4 inches high. The tripod should be located far enough from the tree that each of the four sections of the tree as defined by the poles can be seen in the view finder.

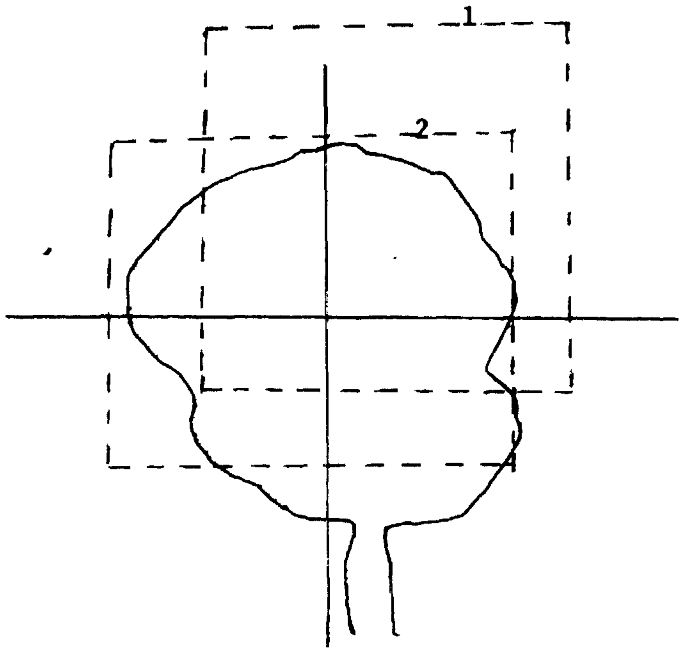
5. Each picture is to be clearly identified by the block, row, and tree number within the block written on a piece of paper at least 8" x 11" in size and displayed so that it is visible in each picture taken but does not hide any of the fruit in the section being photographed. All numbers should be at least 2-1/2 inches high and 1/4 to 3/8 inches thick.

6. Using the settings as determined, take photos of the sections in the following sequence: Lower left, lower right, then move the f-stop 1 stop higher (if it was 8.0 move to 11.0) and take the upper right and upper left portions of the tree. Make sure the tree identification can be seen in all four pictures.



Moving the f-stop one reading adjusts for the increased amount of light in the upper portion of the tree. Also when taking the upper section of a tree, always take as much of the tree as possible making maximum overlap of other sections.

Example



The second coverage is preferable for upper right.

7. The photo record form should be filled out completely.

Row_______________Tree_______________can be filled in ahead of time.

Section entries are:

L.L. - lower left
L.R. - lower right
U.R. - upper right
U.L. - upper left

Film Number - roll and slide number

Shutter Speed $\frac{1}{50}$, $\frac{1}{100}$, $\frac{1}{60}$, $\frac{1}{125}$, etc.

Lens opening f-stop - 5.6, 8.0, 11.0, etc.

Extra slides: If double coverage is taken, identify it by roll and exposure number.

All rolls of film, both the can in which the film comes and the cassettes, are to be numbered consecutively starting with 20. Attempt to expose them in numerical order. If a roll is lost or misplaced before being exposed, use the next roll. If a roll is lost or misplaced after being exposed, retake those trees if possible.

June Fruit Photography Form

Date____________________
Time Started________________
Time Finished_______________

Type Fruit____________________

Block____________________
Variety___________________
Photographer________________
Direction__________________

Row_______________ Tree_______________

Section	Roll and number	Lens setting	Lens speed	Extra slides

Row_______________ Tree_______________

Section	Roll and number	Lens setting	Lens speed	Extra slides

Row_______________ Tree_______________

Section	Roll and number	Lens setting	Lens speed	Extra slides

Row_______________ Tree_______________

Section	Roll and number	Lens setting	Lens speed	Extra slides

Virginia Apples

Time: Start__________
Finish__________

Terminal Limb Fruit Count Form for
Selected Limbs From Itek Prints

Block__________ Row__________ Tree__________

Number of Terminals__________

Objective 1

Cluster	Limb & Path	Fruit Present Yes	Fruit Present No	Cluster	Limb & Path	Fruit Present Yes	Fruit Present No
1				2			

Objective 2

No. or letter	Sect	Count	No. or letter	Sect	Count	No. or letter	Sect	Count	No. or letter	Sect	Count

Check for Missing Terminals (See Instructions)

Objective 3

On Path	Number of Missing Terminals

Record Sheet: Tree Selection and Measurements

Block__________ Date__________ Variety______ No. trees______ No. rows_____

Row interval ($\frac{\text{No. rows}}{8}$)__________ Random Start__________

Tree interval ($\frac{\text{No. tree/row}}{4}$)__________ Random Start__________

Counter__________ Recorder__________ Time: Finish__________ Start__________

Row	Tree	Primary	Rank	Row	Tree	Primary	Rank	Row	Tree	Primary	Rank
Trunk CSA				Trunk CSA				Trunk CSA			
Trunk CSA				Trunk CSA				Trunk CSA			
Trunk CSA				Trunk CSA				Trunk CSA			
Trunk CSA				Trunk CSA				Trunk CSA			
Trunk CSA				Trunk CSA				Trunk CSA			
Trunk CSA				Trunk CSA				Trunk CSA			
Trunk CSA				Trunk CSA				Trunk CSA			

SUPPLIES AND EQUIPMENT FOR 1969 VIRGINIA APPLE PROJECT

1. Ladders (3)
2. Clipboards - 2 per team (one for picture team to have clip bolted to it).
3. Flagging tape - orange and yellow.
4. Fruit count forms - 1 per tree (3 per block) + spares.
5. Photo record form - 1 per block.
6. Battery equiped stereo viewers (2), extra bulbs.
7. Film - 30-36 exp. rolls.
8. Tripod.
9. Camera(s) and light meter.
10. Carpenter aprons.
11. Cardboard tags.
12. Geological survey and county highway maps.
13. Block envelope.
14. Pencils.
15. Magic markers.

Appendix B

Probable Requirements, Costs, and Organization of Future Surveys

General

Not all of the costs incurred on this pilot survey were of the same order of magnitude that might be expected on a larger operational survey. For example, the personnel used in the field were transported from 70 to 90 miles to the sample orchards. Operationally, much of this work would be done by locally hired enumerators. This change should result in both a much lower travel cost and a longer effective work day. Also, the wages of regular supervisory personnel probably would not be as high as for the research statisticians involved with this project. Therefore, the costs listed in this section have been edited to reflect what might be expected on an operational survey. Costs of the various phases of sample selection will be considered separately.

Locating and Identifying Sample Blocks

Qualities required for this job include persistence, a knowledge of the purpose of the survey, the ability to communicate this knowledge, and the ability to both read and draw maps. These qualities can be found (or instilled) in some supervisory enumerators and in some statisticians.

Locating and identifying the sample blocks will take one person about one and one-half hours per block. A major part of this time is spent finding the operator. Increased use of the telephone in setting up appointments could reduce this cost in future surveys. No record was kept of the distance traveled between sample blocks but it probably averaged around 15 to 20 miles per operator--after arriving in the region (Table B-1).

Table B-1.--Cost of locating sample blocks, Virginia 1969

Item	Input	Cost	
		Unit	Total
Vehicle use	17.5 miles/block	$ .10	$1.75
Salary of GS-9	1.5 hours/block	5.03 1/	7.55
Total			$9.30

1/ 1971 GS pay scale.

Selecting, Marking, and Photographing Sample Trees

Selecting and marking the sample trees can be done anytime before the survey period. However, if stereo photographs of the trees are also to be used as a sampling frame for selecting count limbs, this phase of the work must be completed during the winter dormant period. This operation can be accomplished by one well qualified person but a two-person team ordinarily would be preferred.

Qualities required for this job would include the ability to read and make maps, a knowledge of sampling and (if stereo photographs are to be taken) training in taking photographs of the bare trees.

The different operations required of this team would be:

1. Find the sample block.

2. Verify and complete the description and the sketch of the sample block made at the time of the initial contact. The sketch should be on graph paper (1/10 inch squares) so that each square can represent a tree. Missing trees and trees of other varieties should be blacked out on this sketch.

3. Determine the number of rows to be sampled and the number of trees to be selected from each row. The objective would be to distribute the initial sample of trees fairly evenly over the orchard.

4. Measure either the circumference of the trunk or the CSA's of the primary limbs of a systematic sample of trees in a systematic sample of rows. If the measurement is to be used only to array the measured trees by size in preparation for the next phase of sampling, the circumference of the trunk may be adequate. However, if size is to be a covariate in the estimation procedure, the CSA's of the primary limbs should be measured. Cost-wise, the CSA measurements require about an additional two minutes per tree and a special measuring tape. The trees should also be marked in some temporary fashion, such as hanging a piece of flagging tape from a branch.

Table B-2.--Cost of selecting and marking sample trees, Virginia 1969

Item	Input	Cost per unit 1/
Vehicle use	10 miles per block	$1.00 per block
Personnel		
Travel between blocks	20 minutes per block	2.79 per block
Mapping the block	36 minutes per block	5.02 per block
Selecting and measuring 30 to 40 trees	4 minutes per tree	.56 per tree
Marking 9 trees	15 minutes per block plus	2.09 per block
	3 minutes per tree	.43 per tree

1/ The indicated salary costs assume a two-man team of one GS-9 and one GS-5 at 1971 GS pay scale.

5. List the trees in order of size and draw the final sample of trees systematically from this array. This tree would then be marked semi-permanently with paint and with plastic flagging tape wrapped around the trunk.

Selecting Sample Count Units From Photographs

Cost of this phase of the operation would be about as follows:

1. Cost of taking pictures of the tree from opposite sides of the tree - an additional three minutes per tree if done when the sample trees are marked.

2. Cost of film and developing - two slides per tree.

3. Cost of four Itek prints - two work and two field copies.

4. Time for identifying and numbering sample limbs - about one minute per limb.

5. Supervisory and handling time - about 15 minutes per tree.

6. Selection of the sample limbs - ten minutes per tree.

Table B-3.--Cost of selecting sample limbs (s.u.) from stereo photographs, Virginia 1969

Item	Input	Cost per unit 1/
Stereo photography	3 minutes per tree	$ .43 per tree
	2 slides per tree	.64 per tree
Itek prints	4 prints per tree	1.60 per tree
Sample unit		
Delineation	68 minutes per tree plus	3.10 per tree
Selection	5 minutes for each selected sample unit 2/	.23 per s.u.

1/ Wages based on 1971 GS pay scale.
2/ The average number of sample limbs per tree on this survey was 43.4 (Table 13).

Counts of Apples on Sample Limbs

A field crew configuration which worked well in obtaining counts of apples on the pre-selected sample units was one supervisor and four counters. The counters were divided into two-person teams. Each team would work one tree separately and then join the other team on the third and final count tree. The supervisor would make sure the counts were taken on the proper sample units and assist as required. The time required per tree varied from a very few minutes to several hours depending upon the number of the apples and upon the location of the sample limbs in the tree. The average time per sample unit was seven minutes and the average time between trees in the block was five minutes. Detailed cost items are presented in Table B-4.

Table B-4.--Costs of counting apples on sample units, Virginia 1969

Item	Input	Cost
Travel between blocks	10 miles	$2.00 1/
	20 minutes	4.66 2/
Travel between trees in block	5 minutes/tree	.58 3/
Identifying sample limb	3 minutes/s.u.	.35 3/
Counting apples	7 minutes/s.u.	.93 3/

1/ Two vehicles at $ .10 per mile.
2/ One GS-5 supervisor and four GS-3 enumerators at 1971 pay scale.
3/ Two GS-3 enumerators and one-half of a GS-5 supervisor at 1971 pay scale.

Counts of Apples From Photographs

Photographs of the apple trees would be taken by a separate two-person crew (a photographer and an assistant). The average time required to set up, take pictures of one side of a tree, and move to the next tree should not be more than 10 minutes. Taking pictures of both sides of the tree would require, at most, another five minutes.

The cost of getting a count of the apples from one slide—one-fourth of one side of one tree—would include eighteen cents for the film and developing, approximately five minutes for handling and summarization in the office, about ten minutes of interpreter time for each count, and the cost of taking the picture.

The five minutes per slide for handling in the office may seem large--actually, it may understate the actual amount of time required to:

(a) Check in and label each slide,
(b) Determine if the slide is useable,
(c) Distribute the slides to the interpreters for counting,
(d) Collect the slides from the interpreters after counting, and
(e) Sum and record the sum of the interpreters counts.

Table B-5.--Cost of counting apples from photographs, Virginia 1969

Item	Input	Cost
	Cost of acquiring photography	
Travel between blocks	10 miles	$1.00
	20 minutes 1/	2.62 2/
Within blocks		
Between trees	5 minutes 1/	.65 2/
Camera locations per tree	5 minutes 1/	.65 2/
Per slide		
Film and developing	Cost	.18
Handling slides in office	5 minutes 3/	.25 2/
	Training photo interpreters and quality control	
Supervisor	1 hour per interpreter 4/	5.20 2/
Interpreters	2 hours each 3/	5.66 2/
	Counting apples from projected color transparencies	
Interpreter	10 minutes per count 3/	.47 2/

1/ Elapsed time for a two-man crew of a GS-9 photographer and GS-3 assistant.
2/ 1971 GS pay scale.
3/ GS-3 clerk-interpreter.
4/ GS-9 supervisor.

Table B-6.—Summary of anticipated costs, by type of operation and by unit, Virginia 1969

Type of operation	Costs 1/			
	c_1	c_2	c_3	c_4
	Dollars			
Limb counts:				
Tree and sample unit selection	18.11	5.77	.05	
Fruit counts	6.66	.58	1.28	
Size of terminal limb (CSA)	---	---	---	.05
Size of primary limbs (CSA)	2.09	.56	---	---
Counts from photographs:				
Tree selection	---	.43	---	---
Photography	3.62	.65	.65	.18
Interpretation	---	---	---	.72

1/ The different types of costs are defined as follows:

c_1 - costs associated with a particular block, regardless of the size of the sample in the block.

c_2 - costs associated with individual trees, regardless of the sampling scheme in the tree.

c_3 - costs associated with individual sample units for limb counts, or with individual camera locations for photo counts.

c_4 - costs associated with individual slides for photo counts, or in obtaining the size of the terminal limbs (CSA) for limb counts.