A RESEARCH REPORT ON MICHIGAN TART CHERRIES

By

Fred Vogel

Research and Development Branch

Standards and Research Division

Statistical Reporting Service

April 1970

CONTENTS+

	Page
Introduction	1
Summary	1
Data Collection Procedures	2
Analysis	3
I. Analysis of Limb Counts	4
A. Basic Data	4
B. Optimum Sample Allocation	7
C. Comparison of Estimating Procedures	9
II. Analysis of Photo Counts	10
A. Counting Procedure	12
B. Basic Data	13
III. Fruit Weight	17
Appendix	18

Introduction

Objective yield research studies on tart cherries began in Michigan during the 1958 season and continued through 1962. The primary purpose of these studies was to develop workable and efficient field sampling techniques and to obtain measures of variability for constructing sample designs. Major growth and development characteristics were observed in order to determine their relationship with yield per tree.

Research work resumed in 1967 to gain information about various procedures that could be used to improve the efficiency of the sampling methods and increase the precision of the estimates of fruit per tree. One of the main objectives was to develop methods that would reduce, (1) the sample size necessary to attain an acceptable level of precision and (2) the time required to make counts and measurements for each sample tree. The use of photography was investigated with the hope it could be used to:

(1) Select sample limbs independently of the fruit counting phase. This could reduce the time per tree and give a more homogeneous set of sample limbs.

(2) Reduce the sample size necessary to attain estimators with acceptable precision.

(3) Provide fruit counts from photographs to improve estimates for the number of fruit per tree

Summary

This study deals with problems encountered in the early stages of fruit development, since an early season forecast is of primary concern to the industry. Counts of fruit on sample limbs in the selected trees were obtained in 1967 and 1968. Measurements of limb cross sectional area (CSA) were made. Stereo photographs were taken of "bare" trees and color slides of fruit on the trees were taken during the growing season.

Major findings of this study include:

(1) The bare tree stereo slides provided a good basis for constructing a sampling frame for determining the number of terminal sampling units.

(2) The number of fruit on a terminal limb is highly correlated with its cross-sectional area - thus suggesting the use of the CSA in either the sampling procedure or in the estimating procedure.

(3) An optimum allocation indicates that two terminal sampling units per tree be selected. A PPS (sample units selected with probability proportionate to size) estimator showed a sampling variance considerably lower than an equal probability estimator when the sample size is two.

(4) A two stage sampling procedure where one primary is selected with probabilities proportionate to size and two terminal units chosen within the primary, also PPS, performed as well or better than the single stage estimator. This is important from a cost standpoint, because it would only be necessary to obtain limb measurements and counts within one primary rather than in the whole tree.

(5) Two independent estimates of the number of fruit per tree were compared. One was based on a direct expansion of limb counts - the other was obtained from a count of green fruit in the photographs. Analysis of the two estimates show them to be significantly different - thus implying that the counts obtained from the photograph do not reflect the actual fruit count. However, counts of fruit from photographs taken at harvest time do compare favorably with the limb counts.

(6) Measures of tree size - trunk CSA, sum of primary CSA's and area of tree projected on the screen - were compared with total tree production. None showed a significant relationship. This could be due to sample size.

(7) The average weight per cherry at harvest time showed little variation. Thus only a small sample would be needed to obtain a weight estimate of the ripe fruit.

Data Collection Procedures

In June, 1967, two trees were subjectively chosen. A total fruit count by path section and terminal limb (see definitions below) were obtained. The primary purpose of this work was to gain experience dealing with problems to be encountered on future studies.

In April, 1968, eleven trees were selected as follows:

(1) Two conveniently located orchards were chosen in the Belding area of Michigan.

(2) About 200 trees were measured in each block using trunk cross-sectional area (CSA) as a size criterion.

(3) The trees were arrayed by size, and systematic samples of 5 trees from one block and 6 trees from the other block were obtained.

(4) Stereo photographs of the 11 sample trees were taken from opposite sides. On each side, two limbs were tagged with white ribbons and the CSA measured so they could be used as a guide when using photographs to select sample limbs.

The stereo photography was implemented to devise a means of constructing a sampling frame for limb selection. They were used to divide the tree into sampling units (terminal limbs). This procedure is often referred to as "mapping" the tree. The primary objective was to obtain sampling units that were approximately equal in size (about 1.0 CSA). The following definitions were used in the mapping procedure:

(1) Primary limb - all limbs at the first branching of the trunk.

(2) Path section - a section of a primary limb. Either terminal or non-terminal branches emerge from it.

(3) Terminal limb (sampling units) - A limb with a cross-sectional area of about 1.0 square inch. No other major branches should emerge from this limb.

Using the stereo photographs, sketches were drawn of the trees - one sketch per primary. Each primary was then divided into path sections and sampling units. Each sampling unit was numbered in a clockwise manner.

The tree sketches were taken into the orchards in June. CSA measurements were obtained for all sampling units, path sections, and primary limbs. In two trees, a complete count of all of the fruit on the trees was obtained by sampling unit and path section. In the remaining trees, sample limbs were systematically selected using the numbering scheme designated in the mapping process. This hopefully assured a good distribution of terminal limbs around the tree. Fruit on the path associated with each selected limb were also counted.

Photographs - stereo and 35 mm slides - were taken on opposite sides of each tree when the fruit was counted. A metal frame was used to divide each side into quadrants. A separate photograph of each quadrant was taken.

At harvest time, the quadrants were again photographed. Ripe fruit on the sample limbs were picked, counted, and weighed. Counts of fruit visible in the pictures of the different quadrants were obtained for pictures of both green and ripe fruit.

Analysis

The main objective of this study was to determine how to most efficiently estimate the number of fruit in a tree. Different procedures for estimating the fruit, based on a sample of limbs, were compared. Tests were made to determine whether the use of the counts obtained from the photographs could be used to improve the precision of the estimators.

I. Analysis of Limb Counts:

A. Basic Data

To test alternative ways of estimating the number of fruit on a tree, it was necessary to examine some basic tree characteristics and their relationship with the total fruit set. The distribution of the fruit within a tree was also examined.

The first factor considered was the correlation between the size of a sample limb and the number of fruit on it. Two correlation coefficients were computed for each tree sampled in 1968:

(1) The number of terminal fruit vs CSA of terminal limb.

(2) Adjusted terminal fruit vs CSA of terminal limb. Here the fruit on each path section were divided equally among the terminal sampling units on or beyond the path section.

The average size of the sample limbs in each tree was obtained. The standard deviation of the sizes was then expressed as a percentage of the mean (coefficient of variation). This was to test whether the sampling units established were about equal in size. The average number of fruit and its coefficient of variation were also computed. The data is presented in the following tables along with the total number of sample units (N) in each tree and the number sampled (n).

Table I-1.--Tree terminal limb size and fruit counts, Michigan cherries, 1967

Tree	N	Average CSA $\bar{X}$	C.V. (X)	Average fruit $\bar{Y}$ (adjusted)	C.V. (Y)	r (adjusted)
		(Sq. inches)	(Percent)	((Number)	(Percent)	(Correlation)
1	54	1.112	35.6	209.5	58.8	.509**
2	49	1.120	60.4	68.3	64.4	.371**

Table I-2.--Tree terminal limb size and fruit counts, Michigan cherries, June 1968

Tree	N	n	Average CSA (X̄)	C.V. (X)	Average fruit (Ȳ)	C.V. (Y)	Correlation r	Correlation r adjusted
1-6	46	12	1.458	25.6	802.3	33.7	.777**	.736**
2-2	48	5	1.020	42.9	658.4	44.5	.891**	.831**
4-10	31	5	1.400	11.2	591.4	27.1	.801*	.785
10-11	44	6	.900	47.6	526.3	65.0	.924**	.925**
17-17	92	10	.810	24.0	297.4	69.9	.863	.518
22-13	85	11	.863	31.0	304.2	56.3	.829**	.752**
37-1	80	6	.850	46.3	161.8	78.9	.148	.179
42-2	50	10	.770	26.6	378.4	66.3	.670*	.796**
42-15	74	10	.710	37.7	373.7	68.0	.578	.581
11-13	22	22	.909	43.3	599.5	53.3	.867**	.850**
11-6	28	28	1.017	45.8	301.2	72.1	.680**	.631**
Average			.988	42.2	446.4	68.1	.680**	.690**

* Significant at 95% probability level

** Significant at 99% probability level

The data indicates there was some variation in the average size of a sampling unit even though an attempt was made to equalize them. Since the designation of sample limbs was based on photography alone, the amount of variability is not very large. The bare tree photography thus provided a good frame from which to select sample limbs.

The correlation coefficients revealed a significant relationship between the size (CSA) and number of fruit on a terminal limb. There is only a slight difference in the correlation between the adjusted and actual fruit counts and the terminal limb size measurement. Therefore, if a PPS method of sample selection is used, little loss in efficiency will result by allocating path fruit to terminal limbs.

Table I-3 shows the correlation coefficients between some additional measures of tree size and numbers of tree fruit. The data for all trees were combined to obtain these correlations.

Table I-3.--Relationship between trunk and primary limb sizes and tree fruit set, Michigan cherries, June 1968

Relationship	Degrees of freedom	Correlation coefficient
Primary limb CSA vs. number of terminals on the primary	51	.727**
Primary limb CSA vs. estimated number of fruit on the primary	51	.616**
Trunk CSA vs. estimated number of fruit on the tree	10	.422
Sum of primary CSA's within each tree vs. estimated number of fruit on the tree	10	.300
Total area of tree as measured from photo vs. estimated fruit on the tree	10	.500

** Significant at the 99% probability level.

The primary limb CSA is highly correlated with both the number of fruit it contains and the number of terminal sampling units in it.

The significance of the correlation coefficient between terminal limb CSA and terminal fruit suggests that the sampling design should utilize this size criterion. Since CSA's at the primary level are still significantly related to fruit set, this size measurement can also be utilized. These factors suggest the use of a two-stage sampling procedure using limb CSA's in each stage to estimate the number of fruit per tree.

Before comparing the relative efficiencies of some sampling designs, the distribution of the fruit within a tree will be examined. This is done in the following section and is used to determine optimum sampling fractions.

B. Optimum Sample Allocation

The next problem to be considered is the optimum sample allocation. Of primary concern is the optimum sampling fractions to use when sampling within a tree. The optimum sampling rate for the number of trees to select within a block will also be determined. The cost and variance functions given below will be considered.

$$\text{Total cost} = C_1 n_1 + C_2 n_1 n_2 + C_3 n_1 n_2 n_3 + C_4 n_1 n_2 n_3 n_4$$

$$\text{Var (Y)} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2} + \frac{\sigma_3^2}{n_1 n_2 n_3} + \frac{\sigma_4^2}{n_1 n_2 n_3 n_4}$$

C_1 = The amount of travel time between blocks.

C_2 = The time period involved in finding a sample tree and identifying all primary limbs in the tree.

C_3 = The number of minutes it takes to select one primary and determine the number of terminal sampling units in it.

C_4 = The number of minutes it takes to select a terminal unit, identify and count the number of fruit it contains.

Let n_1 = number of blocks

n_2 = optimum number of trees to sample in a block

n_3 = optimum number of primaries to sample in a tree

n_4 = optimum number of terminals to sample in a primary

When using the above variance and cost functions, Snedecor and Cochran (1967) 1/ show that the following values of n_i are optimum:

1/ Snedecor, George W. and William B. Cochran, Statistical Methods, Sixth Edition, Iowa State University Press, Ames, Iowa, 1967 pp 531-534.

$$n_4 = \sqrt{\frac{C_3\ S_4^2}{C_4\ S_3^2}} \qquad n_3 = \sqrt{\frac{C_2\ S_3^2}{C_3\ S_2^2}} \qquad n_2 = \sqrt{\frac{C_1\ S_2^2}{C_2\ S_1^2}}$$

Variance components for each level of sampling were computed. The data was from the sample limbs in the eleven trees used in 1968. Since the sample limbs were chosen with equal probabilities - the allocation will be optimum for a simple random sampling scheme.

Table 14.--Cost data and variance components, Michigan cherries, June 1968

Component	Cost	d.f.	Variance
Between blocks	30 minutes	1	11,314,529
Trees within blocks	15 minutes	9	76,897,354
Primaries within trees	20 minutes	39	14,510,890
Terminals within primaries	30 minutes	74	132,292,438

The variance components were obtained from the nested AOV based on expanded numbers of fruit:

The optimum sampling rates for equal probability sampling follow:

$n_4 = 2.4 = 2$ or 3

$n_3 = .4 = 1$

$n_2 = 3.7 = 4$

Since it will probably be necessary to minimize variance subject to a fixed cost rather than minimizing cost subject to a fixed variance, n_i will be determined from:

$$\text{Total cost} = n_1\ (C_2\ n_2 + C_3\ n_2\ n_3 + C_4\ n_2\ n_3\ n_4).$$

The evidence is fairly conclusive that only one primary unit per tree be selected. The optimum number of terminals to select within a primary is two or three. Since the between block component of variance has only one degree of freedom, the value of n_2 may be questionable. However, Small's study (1967) 1/ based on six blocks located in three separate districts, also gives an optimum allocation of 2 or 3 trees per block.

1/ Small, Richard P., Research Report on Tart Cherry Objective Yield Surveys, USDA, Statistical Reporting Service Research Publication distributed December, 1967.

C. Comparison of Estimating Procedures:

Consider a tree consisting of a population of terminal limbs, each containing an unknown number of fruit. The objective is to find a practical sampling procedure that provides a good estimate of the total number of fruit in the tree.

The analysis to this point has shown the following results:

(1) Terminal limb CSA's are correlated with numbers of fruit.

(2) Primary limb CSA's and fruit numbers are correlated.

(3) When considering costs - two terminals from the same primary should be sampled when selecting limbs with equal probabilities.

A total fruit count - by path section and terminal sampling unit - was available for four trees. These trees were completely "mapped" making it possible to compare several sampling designs.

A single stage sampling scheme is donsidered first. This method requires a listing of all of the terminal limbs within each tree. This listing is prepared so the selection of a terminal limb does not depend upon its associated primary limb.

Two procedures for drawing a sample from this listing were considered.

(1) Select the terminal limbs with equal probabilities.

(2) Select the sample with probabilities proportional to size (CSA). With PPS sampling, the listing must also contain a size measurement for each terminal limb.

Previous analysis has suggested a PPS sample be selected. However, the equal probability scheme is presented to provide a means for comparison. Sample sizes of $n = 1$ and $n = 2$ are compared. Population coefficients of variation (C.V.) for the following single stage sampling schemes are shown in Table I-5.

$\hat{Y}_1$ - one unit chosen with equal probabilities (EP).

$\hat{Y}_2$ - one unit chosen with probabilities proportional to size (PPS).

$\hat{Y}_3$ - two units selected (EP) without replacement.

$\hat{Y}_4$ - two units selected (PPS) without replacement.

The single-stage sampling schemes become impractical when the number of terminals in a tree is large due to the cost of preparing the listing.

An alternate procedure would be to consider two-stage sampling schemes. Since the optimum allocation indicates only two terminals be selected from one primary, different procedures for selecting two terminals from one primary will be compared.

As the name implies, the sampling is done in two steps. First, a primary limb is selected. Then the sample of terminal limbs is drawn from this primary. This method has some distinct advantages. For example, it is not necessary to know the number and size of every terminal limb in the tree. Only the units within the selected primaries need be identified and measured. A disadvantage is that sampling variances are usually larger than those from a single stage scheme because there is sampling error at two stages rather than one.

The two-stage sampling procedures considered are:

$\hat{Y}_5$ - one primary chosen with equal probabilities, two terminals within the selected primary chosen with equal probabilities. This was suggested by the optimum allocation.

$\hat{Y}_6$ - one primary (PPS), one terminal within (PPS).

$\hat{Y}_7$ - one primary (PPS), two terminals within (PPS),

The variance estimator for the two-stage procedure also gives the variance in components - one due to sampling primaries within a tree, and another due to sampling terminals within primaries.

Another selection procedure considered is known as the "Random Path" 1/ method. This method consists of proceeding out a selected primary and stopping at each branching point to determine which one to follow. A PPS procedure is used at each branching point. The path is followed until a terminal unit of a desired size is reached.

The sampling variance for each estimator was computed to provide a means for comparison. The formulas used in each case are presented in Appendix A. The coefficients of variation for the various estimators for the four trees completely counted are shown in Table I-5.

The effects of PPS sampling as compared to equal probability sampling are most noticeable in the single-stage case. An encouraging note is the performance of the two-stage sampling schemes. This indicates

---

1/ Jessen, Raymond J. (1955). Determining The Fruit Count on a Tree by Randomized Branch Sampling. Biometrics.

Table I-5.--Coefficients of variation for estimators, Michigan cherries, 1967 and 1968

Estimator	Tree 1 N = 54	Tree 2 N = 49	Tree II-6 N = 28	Tree II-13 N = 22
Single stage				
$\hat{Y}_1$ EP n = 1	.61	.62	.72	.52
$\hat{Y}_2$ PPS n = 1	.44	.70	.49	.25
$\hat{Y}_3$ EP n = 2	.41	.44	.50	.36
$\hat{Y}_4$ PPS n = 2	.33	.49	.35	.17
Two stage				
$\hat{Y}_5$ one prim. EP two term. EP	.53	.54	.52	.36
$\hat{Y}_6$ one prim. PPS one term. PPS	.49	.62	.51	.26
$\hat{Y}_7$ one prim. PPS two term. PPS	.36	.47	.41	.19
Random path				
$\hat{Y}_8$ n = 1	.57	.80	.59	.39
Number of fruit in trees	11,311	3352	8364	13,250

little or no loss in sampling efficiency will occur when sampling in two stages. Furthermore, considerable time will be saved if it is necessary to "map" only one primary limb in a tree.

The estimator designated by Y_5 (selection of one primary with equal probabilities and two terminals also with equal probabilities) is that suggested by the optimum allocation. The PPS sampling schemes (Y_6 and Y_7) provide two alternatives that are superior to the optimum allocation:

(1) If $\hat{Y}_6$ is used, then we need to sample only one terminal limb instead of two to get about the same sampling error as that for the optimum allocation.

(2) If $\hat{Y}_7$ is used, a considerable reduction in sampling errors is achieved for the same sample size.

The variances of the two-stage estimator consist of two components - one due to primaries and one due to sampling terminals within primaries. Table I-6 shows how these components vary for different sampling schemes and sample sizes.

Table I-6.--Two-stage variance components, Michigan cherries, 1967 and 1968

Estimator	Primaries	Terminals	Total	Primaries	Terminals	Total
	Tree 1			Tree 2		
$\hat{Y}_5$	14,160,170 +	21,253,634 =	35,413,804	813,604 +	2,419,392 =	3,232,996
$\hat{Y}_6$	3,161,413 +	27,221,407 =	30,382,820	557,121 +	4,115,825 =	4,672,946
$\hat{Y}_7$	3,161,413 +	13,340,664 =	16,502,077	557,121 +	1,948,741 =	2,505,862
	Tree 11-16			Tree 11-13		
$\hat{Y}_5$	15,305,375 +	8,736,695 =	24,042,070	8,470,803 +	13,690,485 =	22,161,288
$\hat{Y}_6$	7,024,609 +	10,989,196 =	18,013,805	2,212,857 +	9,564,051 =	11,776,908
$\hat{Y}_7$	7,024,609 +	4,796,732 =	11,821,341	2,212,857 +	3,987,867 =	6,200,724

II. Analysis of Photo Counts

A. Counting Procedure:

The goal of this analysis was to obtain knowledge of the possible advantages of photography in improving estimates of fruit production. This section discusses the procedures followed and conclusions reached.

When limb counts and measurements were obtained in June, 1968, color photographs of the trees were taken. A metal frame was used to divide each side of each tree into quadrants - thus a total of eight pictures were taken of each tree. Similar pictures were again taken at harvest time.

Several photographs were used to train the counters. As expected it took a considerable amount of time to count an entire slide. It also appeared that two different people counting the same slide might obtain significantly different counts. Based on the experience gained during the training period, the counting procedure developed for the green fruit slides was as follows:

(1) Four out of the eight photographs in each tree were selected - two quadrants comprising a diagonal from each side.

(2) Four people did the actual counting.

(3) Within each tree, each counter was randomly assigned to a quadrant. Then in order to replicate the counts so any differences between counters could be detected, they were again assigned to the quadrants, but subject to two restrictions:

   (a) No photo was to be counted twice by the same person.

   (b) Each person was to count once on each side of each tree.

(4) Each photo-slide was projected onto a screen marked off into square grids. The grids provided a sampling frame so a subsample of columns or grids within columns could be selected.

(5) The first person assigned to a quadrant counted it entirely, recording the counts grid by grid. The second person assigned to the quadrant counted only the fruit in a few systematically selected columns. A sampling fraction of one third was used.

This counting design provided a replication of columns within each quadrant and a means for detecting any differences between counters. A column by column comparison will test the feasibility of subsampling within a quadrant.

A similar procedure was followed when counting the fruit in the pictures of ripe cherries. The main difference was that only two counters were used. They both counted only a subsample of each slide.

B. Basic Data

Selected columns within each slide were counted twice - each count by a different person. Column by column comparisons thus provided a test to determine if there were any differences between counters. No significant differences between counters were found. The absence of any differences between counters indicates it would not be necessary to replicate the counts of every photograph in an operational program. It may still be advisable to replicate a portion of the counts as a check on the counters.

Table II-1 shows the average time in minutes it took to count an entire slide and to count a fraction of a slide. The sampling procedure followed cut the average time in half. The comparisons of counting times between green and ripe fruit indicates ripe fruit was much easier to count.

Table II-1.--Average number of minutes to count selected photographs, Michigan cherries, 1968

Tree	Green fruit			Ripe fruit	
	Estimated number of fruit	Averge time to count entire photo	Average time to sample photo	Estimated fruit	Average time to sample photo
		Minutes	Minutes		Minutes
1-6	36,899	40	18	30,073	7
2-2	31,603	31	22	26,325	5
4-10	18,339	28	13	16,597	8
10-11	23,158	44	27	22,741	8
11-6	8,364	25	13	7,360	4
11-13	13,250	41	11	10,626	4
17-17	27,370	48	22	23,292	5
22-13	25,859	62	27	24,307	6
37-1	12,960	32	15	11,664	5
42-2	18,935	41	23	14,334	5
42-15	27,661	42	21	25,254	6
Average		39.5	19.3		5.7

In order to utilize the photography on a large scale, it will be necessary to reduce the number of slides per tree and count only a portion of a slide.

Variance components were computed for the levels of sub-sampling within a tree. These were used to determine the feasibility of subsampling. The costs incurred at each level were also considered. They were:

C_1 - Cost of moving from tree to tree.

C_2 - Cost per photograph within a tree - includes film and developing charges.

C_3 - Time involved in determining what columns to sample.

C_4 - Average time to count the fruit in a sample grid.

The following table gives the cost data and variance components for the levels of subsampling.

Table II-2.--Cost data and variance components, Michigan cherries, June, 1968 data

Subsampling level	Cost (dollars)	d.f.	Variance component
Between trees	2.00	10	2.66
Photographs within a tree	.30	33	6.53
Columns within a photo	.05	455	4.53
Grids within a column	.05	3154	26.12

Following the procedure outlined in section I-B the optimum allocation is:

(1) Three grids per sample column should be selected.

(2) Two columns within a slide would be chosen.

(3) Four quadrants (photos) are needed from each tree.

The number of trees to photograph would depend upon the size of the budget.

The estimated total number of fruit in each tree was obtained by multiplying the average number of fruit per quadrant by the number of quadrants in each tree. An independent estimate of the fruit set in each tree was available from the limb counts made at the time the pictures of green fruit were taken. Limb counts were also obtained with the pictures of ripe fruit. These estimates of fruit set are shown in Table II-3. Table II-3 gives the correlation coefficients between the two estimates.

The difference between the two estimates based on limb counts reflect the fruit drop occurring during the growing season.

Despite a decrease in fruit set, more fruit were counted from the photos of ripe fruit than from the green fruit photos. The correlation coefficient indicates the number of green fruit visible in a photograph is not related to the actual number of fruit in the tree.

Table II-2.--Estimated total number of fruit in trees by photo and limb counts, Michigan cherries, 1968

Tree	Green fruit			Ripe fruit		
	Limb Counts	Photo counts	Photo % of limb	Limb counts	Photo counts	Photo % of ripe limb
1-6	36,899	4010	11	30,073	8,818	29
2-2	31,603	3464	11	26,325	5,452	21
4-10	18,339	2323	13	16,597	3,732	22
10-11	23,158	4680	20	22,741	8,650	38
11-6	8,364	1399	17	7,360	2,134	30
11-13	13,250	3191	24	10,626	3,162	30
17-17	27,370	7031	26	23,292	8,492	36
22-13	25,859	7723	30	24,307	10,092	42
37-1	12,960	3101	24	11,664	2,704	23
42-2	18,935	3232	17	14,334	7,526	52
42-15	27,661	3893	14	25,254	10,456	41
Average		r = .56	18		r = .87	34

Two factors may contribute to these problems:

(1) The fruit set in the tree may be such that not all fruit are visible in the photograph. The green fruit may also blend with the leaves.

(2) The quality of the photographs could be improved. A polaroid lens filter has been suggested to reduce the glare and shading problems. Also, using a faster film then Kodachrome II (ASA rating of 25) would permit using smaller lens openings and greater depth of field.

It appears that improved photographic techniques will be needed if it is to be utilized at the green fruit growth stage. The lack of correlation between the green photo and limb counts cannot be caused by a counter error. The comparisons between counters showed that they were counting the same number of fruit in a photograph.

The correlation between the ripe photo counts and limb counts demonstrates the feasibility of using the photographs. However, these were taken at harvest time. The cherry industry is primarily interested in an early season forecast. Any further attempts to utilize the photography should be directed toward the fruit in its early stages of growth.

III. Fruit weight

During harvest time in 1968, fruit from about three terminals in each tree was picked and weighed. AN AOV was computed to determine if there are any differences within trees, between trees, or between blocks.

Table III-1.--Nested AOV, ripe fruit weight, Michigan cherries, 1968

Source	D.F.	Sum of squares	Mean squares	F
Blocks	1	.20	.20	.61
Trees/blocks	9	3.02	.34	.66
Terminals/tree	21	10.67	.51	

This table supports the findings of earlier studies. There is no significant difference in the average weight of fruit on terminals within each tree. In other words, the average weight per cherry on one terminal limb can be expected to be the same as the average weight of those on other terminals within a tree. Furthermore, the average weight per fruit on one tree can be expected to be the same as that on another tree within a block. The table also shows no significant difference between the block means, but with only two blocks to compare, no general inferences should be made.

The uniformity of the weight data indicates that a relatively small sample as compared to that required for count data is necessary to estimate the average weight per cherry.

Appendix A - Estimating Procedures

The following terms will be used:

N = The number of primary limbs in the tree.

M_i = The number of terminal sampling units in the i^{th} primary.

X_i = The cross-sectional area (CSA) of the i^{th} primary.

$P_i = \frac{X_i}{\sum^N X_i}$ = The probability that the i^{th} primary limb was selected.

X_{ij} = The CSA of the jth terminal unit in the i^{th} primary unit.

$T_{ij} = \frac{X_{ij}}{\sum_j^{M_i} X_{ij}}$ = The probability that the j^{th} terminal unit in the i^{th} primary limb was selected.

Y_{ij} = The number of fruit associated with the j^{th} terminal in the i^{th} primary.

Then the total number of fruit in the tree is $Y_{..} = \sum_i^N \sum_j^{M_i} y_{ij}$, and the number of sampling units is $\sum_i^N M_i$. Also $\sum_j^{M_i} y_{ij} = Y_{i.}$ is the number of fruit on the i^{th} primary unit.

$$\bar{Y} = \frac{\sum_i^N Y_i}{N} \text{ while } \bar{\bar{Y}} = \frac{\sum_i^N \sum_j^{M_i} y_{ij}}{\sum M_i}$$

Now we have a complete count of the total number of fruit in the tree by terminal unit. This will be used to compare several estimators that could be used to estimate the number of fruit in the tree. The methods to be considered are - single stage, two stage, and random path.

I. Single stage

a. Select one terminal unit with equal probabilities of selection.

Then $\hat{Y}_1 = (\sum_i^N M_i)\, y_{ij}$ and

$$\text{Var}\ (\hat{Y}_1) = \frac{(\sum^{N} M_i)^2 \quad (\sum M_i - 1)}{\sum M_i} \quad \frac{\left[\sum_i^N \sum_j^{M_i} \quad (y_{ij} - \bar{\bar{Y}})^2\right]}{\sum M_{i - 1}}$$

b. Select one terminal sampling unit with probabilities proportional to a measure of size (CSA). The probability of any terminal unit being selected is

$$z_{ij} = \frac{X_{ij}}{\sum_i^N \sum_j^{M_i} X_{ij}}$$

Then $\hat{Y}_2 = \frac{y_{ij}}{z_{ij}}$ and

$$\text{Var}\ (\hat{Y}_2) = \sum_i^N \sum_j^{M_i} \quad z_{ij} \quad \left(\frac{y_{ij}}{z_{ij}} - Y_{..}\right)^2 = \sum_i^N \sum_j^{M_i} \quad \frac{y_{ij}^2}{z_{ij}} - Y_{..}^2$$

Where Y.. is the total fruit in the tree.

c. Select two terminals with equal probabilities.

$$\hat{Y}_3 = \left[\sum_i^N M_i\right] \frac{(y_{ij} + y_{ij'})}{2}$$ and

$$\text{Var}\ (Y_3) = \frac{\left[\sum_i^N M_i\right]^2 \quad (\sum_i^N M_i - 2)}{(\sum^N M_i) \quad (2)} \quad \frac{(\sum_i^N \sum_j^{M_i} \quad (y_{ij} - \bar{Y})^2}{\sum_i^N M_i - 1}$$

d. Select two terminals with unequal probabilities (PPS). The probability of any terminal unit being selected first is

$$Z_{ij} = \frac{X_{ij}}{\sum_i^N \sum_j^{M_i} X_{ij}} = Z_k$$ which is similar to that for Y_2.

$$\hat{Y}_4 = \frac{1}{(2 - Z_k - Z_{k'})} \left(\frac{y_k}{Z_k} (1 - Z_{k'}) + \frac{y_k}{Z_{k'}} (1 - Z_k) \right)$$

$$\text{Var } (\hat{Y}_4) = \sum_{K\,K'}^{M_i} Z_k Z_{k'} \frac{(1 - Z_k - Z_{k'})}{(2 - Z_k - Z_{k'})} \left(\frac{y_k}{Z_k} - \frac{y_{k'}}{Z_{k'}} \right)^2$$

II. Two-stage Sampling

a. Suppose that one primary unit was selected with equal probabilities. Then two terminals were selected with equal probabilities within the chosen primary. An unbiased estimator is

$$\hat{Y}_5 = \frac{N M_i (y_{ij} + y_{ij})}{2}$$

$$\text{Var } (\hat{Y}_5) = N \Sigma Y_i - Y)^2 + N \sum_i^N M_i \frac{(M_i - 2)}{2} \sum_j^{M_i} \frac{(y_{ij} - \bar{y}_i)^2}{M_{ij} - 1}$$

Notice that the variance of the estimate is in two components. -

(1) Represents the variation among the primary totals Y_i.

(2) Represents the variation among the terminal units within each primary.

Select one primary with probabilities proportional to size (C.S.A.) then select one terminal within the primary again with PPS.

$\hat{Y}_6 = \frac{y_{ij}}{P_i \, t_{ij}}$ is an unbiased estimator of the population total and

$$\text{Var } (\hat{Y}) = \sum_i^N \frac{Y_{i.}^2}{P_i} - Y_{..}^2 + \sum^N \frac{1}{P_i} \left(\sum_j^{M_i} \frac{y_{ij}^2}{t_{ij}} - Y_{i.}^2 \right)$$

c. Select one primary with probabilities proportional to size. Then select two terminals from within the selected primary (wtr).

$$\hat{Y}_7 = \frac{1}{P_i} \cdot \frac{1}{(2 - T_{ij} - T_{ij'})} \left[\frac{y_{ij}}{T_{ij}} (1 - T_{ij'}) + \frac{y_{ij'}}{T_{ij'}} (1 - T_{ij}) \right]$$

$$\mathrm{Var}\,(\hat{Y}_7) = \sum_i^N \frac{Y_i^2}{P_i} - Y_{..}^2 + \sum_i^N \frac{1}{P_i} \sum_{j\,j'}^{M_i} \frac{T_{ij}\, T_{ij'}\, (1 - T_{ij} - T_{ij'})}{(2 - T_{ij} - T_{ij'})} \left(\frac{y_{ij}}{T_{ij}} - \frac{y_{ij'}}{T_{ij'}} \right)^2$$

III. <u>Random Path</u>

Select a terminal limb by the random path method as described by Jessen. Calling the final probability of selection R_{ij}, the unbiased estimator of Y_8 is

$$\frac{y_{ij}}{R_{ij}} .$$

He suggests using the following variance estimator -

$$\mathrm{Var}\,(Y) = \sum_i^N \sum_j^{M_i} R_{ij} \left(\frac{y_{ij}}{R_{ij}} - Y_{..}\right)^2 = \sum_i^N \sum_j^{M_i} \frac{y_{ij}^2}{R_{ij}} - Y_{..}^2$$