Methods Used in a Survey of Orchards (Michigan Cherries)

By H. F. Huddleston

Comparisons are here made of different methods of estimating the number of cherry trees in a survey of commercial orchards. The author also discusses some problems encountered in the use of lists and gives a comparison of costs when the mailed questionnaire and the interview sample of non-respondents are used.

IN MICHIGAN the cherry industry is continuing to grow in importance. Every alternate year since 1940 has brought a new record crop and in the last 5 years production was 44 percent larger than it was in the preceding 5 years. With this expansion many questions have arisen in the minds of both growers and processors. The Michigan Cooperative Crop Reporting Service, representing BAE and the Michigan Department of Agriculture, was requested to get answers to the following questions to serve as a long-range planning guide to growers and processors:

1. Is the increase in production due to an increase in plantings, to trees of a longer life, to an increase in yield per tree, or to a combination of any or all of these factors?
2. If there has been an expansion in tree numbers, where has it taken place?
3. What is the age distribution of present trees?
4. What number of bearing trees should be used when estimating production in the next few years—until 1950 census data are available?

A survey was made with RMA funds. The industry wanted data on tree numbers by counties and more detailed information by crop-reporting districts. The funds, personnel, and time available indicated a combined mail-and-interview survey as the most efficient design.

In this survey a commercial cherry orchard (or farm) was defined as an orchard with 50 or more cherry trees. However, an orchard (or farm) in this case represents an operating unit and includes, in some instances, two or more farms. For example, if a father and son operated two farm tracts jointly, these tracts were considered as one unit.

Questionnaires were mailed during the third week of February 1949 and a follow-up mailing to non-respondents was made in early March. Usable returned schedules totaled 2,170 or 54 percent of those mailed out. To learn whether the operations of those producers who did not respond differed significantly from those who replied, 20 names of non-respondents, plus several alternates, were drawn at random in each of 10 counties. These orchardists were interviewed between March 22 and April 7. This technique of integrating mail and interview sampling was used in arriving at estimates for counties. In addition, district production data were used as a check on tree numbers. The interviewing was done in the following counties: Antrim, Grand Traverse, Leelanau, Benzie, and Manistee in the northwest district; Mason and Oceana in the central west district; Allegan, Van Buren, and Berrien in the southwest district.

The number of growers interviewed was proportional to the number of non-respondents in the stratum or district, with the restriction that 20 growers were to be interviewed in each county selected. The probability of any county in a stratum being selected was proportional to the number of non-respondents in the county.

There are several sources of error in a survey of this kind. First, it is almost impossible to obtain a complete list of current growers because of changes in ownership or rentals, or because an orchard may be listed twice under different names. Second, there may be errors, for example, in memory; some growers do not know how many trees they have or when they were planted. Third, the results are subject to sampling errors. In general, the sampling error for total district tree numbers, including both bearing and non-bearing, can be expected to be less than 10 percent in 19 cases out of 20, as explained later. The greatest sampling

errors are to be expected where relatively few growers reported on items, such as the percentage of fall plantings in the large size groups.

For the interview sample of 200 farms, the total travel and personnel costs were $2.71 per schedule. Enumerators averaged 7.5 interviews per day and traveled 13.5 miles per interview. Total costs for the interview sample were about $4 per schedule. Total costs for the 2,170 schedules obtained by mail were about 70 cents a schedule.

Some Characteristics of Interview Sample and Mail Returns

The sample of non-respondents for the 10 counties was studied to gain knowledge of possible mail biases. Those points which seemed to be of general interest were selected for examination. Some differences between the mail respondents and the non-respondent universe were noted, but they are not unlike those frequently found in other populations sampled for agricultural data. The following differences were clear: (1) The mail respondents are dominated by the larger growers; (2) the mail respondents have been planting more trees during recent years, as evidenced by the smaller percentage of bearing trees (hence they had a greater interest in the survey); and (3) the mail respondents apparently have a smaller proportion of sweet-cherry trees in their orchards (except in the southwest district), but some growers may have failed to report sweet-cherry trees because, comparatively, they are not commercially important.

Successive Mailings in Grand Traverse and Leelanau Counties

Selectivity in mail returns has long been one of the major weaknesses in the voluntary mail sample, particularly in special surveys for which no historical series are available with which to true up the results. Third and fourth mailings were made to Grand Traverse and Leelanau Counties, in the hope of getting some information about a possible bias in the mail return and about possible duplications in the mailing list.

Several points regarding method are obvious from a study of the data. The response dropped off sharply after returns from the second request were in. The trend in returns from successive mailings was not consistent at the county level, but when the counties were combined the trend was more consistent. This seems to bear out the preliminary conclusion of Hendricks[1] that adjustment for bias from such a trend in a mail survey may be of limited usefulness in a small finite population; probably 50 to 100 returns are needed from successive mailings to learn consistent trends. Publicity by radio, newspapers, and local Extension people was stopped after the second request was made. Hence the response to the third and fourth requests was rather low. But the two additional mailings to Grand Traverse and Leelanau Counties were of most value in verifying the presence of duplications in the mailing lists and for estimating the number of them. Approximately one-third of those who responded to the third and fourth requests indicated that their orchards had reported previously under a different name.

District and County Estimates

District estimates of the total number of sour-cherry trees were obtained as follows: A ratio estimate[2] (the total production in each district for non-respondents, multiplied by the ratio of trees to production in the interview sample) was computed; this figure was added to the number of sour-cherry trees for the stratum reported by mail. The integrated estimate is thus derived from both the mail returns and the interview sample of non-respondents. This total production for all non-respondents in a district was obtained by taking the difference between the total production for the district—as estimated by the Agricultural Statistician—and the production reported by the growers who answered the mail inquiry.

An alternative estimating procedure, a regression or double-sampling method, was also tried for each district.[3]

First, an estimate of tree numbers was derived for each county by multiplying the average trees per grower, as reported on the mail questionnaire,

[1] HENDRICKS, WALTER A. ADJUSTMENT FOR BIAS BY NON-RESPONSE IN MAILED SURVEYS. This magazine 1 (2): 52–56, 1950.

[2] COCHRAN, W. G. SAMPLE SURVEY TECHNIQUES. N. C. State College and Bur. Agr. Econ., 1948. (Processed.)

[3] FINKNER, A. L. METHODS OF SAMPLING FOR ESTIMATING COMMERCIAL PEACH PRODUCTION IN NORTH CAROLINA, N. C. Agr. Expt. Sta. Tech. Bul. 91. 1950.

TABLE 1.—*Estimated tree numbers from mail returns alone and from mail returns plus non-respondent interviews for counties in which non-respondents were interviewed*

County	Mail estimate	Integrated estimate (mail and interview consolidated)	1945 census
	Thousands	*Thousands*	*Thousands*
Antrim	121	108	101
Benzie	164	143	147
Grand Traverse	787	715	572
Leelanau	553	498	418
Manistee	95	87	76
Mason	119	143	99
Oceana	558	529	521
Allegan	95	78	83
Berrien	391	340	330
Van Buren	214	197	121

by the number of growers per county. The only control data, or basis for expansion into county estimates, was the number of growers per county. A second estimate was derived for each of the 10 counties with an enumeration sample as follows: The average number of trees per orchard, reported in the sample of 20 enumerated orchards, was multiplied by the number of growers who did not reply to the mail inquiry. This estimate for the non-respondents was added to the number of trees reported by mail to derive an integrated, and presumably an unbiased, estimate of all trees in the county. These two sets of estimates are shown in table 1, together with 1945 census data, for purposes of comparison.

For the 10 counties in which an enumeration was made, a regression equation was set up, using the estimates from the mail survey as the independent variable (X) and the unbiased estimate as the dependent variable (Y). Estimates for the 14 counties without enumeration samples were derived from this formula by substituting the mail-sample estimates for "X."

In a few of the smaller counties the regression equation did not seem suitable because of the magnitude of the Y-intercept. That is, a negative Y-intercept resulted in a computed number of trees for the non-respondents so low that it looked unreasonable. In these counties a per orchard expansion, derived from the mailed returns, was used to estimate tree numbers for those orchards for which neither county agents nor inspection records could supply data. But in the major counties the regression estimate was accepted.

Our list of growers was incomplete for most counties in the northwest district and for some counties in the central west district. Tree numbers for the missing growers were obtained from data collected in connection with the 1948 cherry fruit-fly inspections. Comparing district totals for the two methods it was found that the latter were considerably larger in the northwest and central west districts. The two methods gave about the same results in the southwest district. The difference between the per orchard estimate and the ratio estimate in the northwest and central west districts was attributed to duplications in the list of operating units. The most tenable theory seemed to be that the duplication was approximately proportional to the product of the number of respondents to the mail survey in the county multiplied by the average number of trees per grower in the county.[4] Consequently, we formulated the hypothesis that the remaining duplication of operating units would be proportional to the product of the number of non-respondents and the average number of trees per grower in the county.

An alternative hypothesis that the duplication was a linear function of the number of non-respondents seemed as good within districts. However, if several districts were to be taken together or if the State as a whole were to be considered, the first hypothesis seemed preferable and simpler. The magnitude of the duplication was made evident by comparing the total number of growers in the county with the number of orchards in the 1945 Census of Agriculture in the northwest district, where virtually all cherry orchards are commercial units. When this comparison was made it was found that in several counties the number of growers on our list was considerably in excess of the census number of orchards; in fact, the excess of grower names over the number of orchards in Grand Traverse County, as given by the census, supported the above hypothesis which was used to eliminate the duplications in the larger counties. A ratio estimate of sweet-cherry tree numbers could not be computed, as reliable pro-

[4] A larger orchard is more likely to be duplicated than a small orchard since several operators or owners or both may be associated with the transactions of the orchard; also the larger the number of orchards in a county the larger the number of duplications.

TABLE 2.—*District and State data for all cherries, showing magnitude of adjustments applied to original regression estimates*

District and State	1 Regression estimate	2 Missing growers from insp. data	3 Per orchard estimate (1) + (2)	4 Indicated duplication (Trees)	5 Duplication removed (Trees)	6 Final estimate (3) − (5)	7 1945 census All cherries
	Thousands	*Thousands*	*Thousands*	*Thousands*	*Thousands*	*Thousands*	*Thousands*
Northwest	1, 619	299	1, 918	331	[1] 326	1, 592	1, 360
Central West	733	162	896	105	[1] 96	800	699
Southwest	646	0	646	36	36	610	533
Remainder of State	58	0	58	0	[2] 4	54	211
State	3, 056	461	3, 518	472	462	3, 056	2, 803

[1] Duplication not removed in several counties because of suspected compensating bias in mail survey, that is, growers forgot to report sweet cherry trees; in several counties it was possible to remove duplication by direct comparison rather than according to the stated hypothesis.

[2] Column 5 was obtained by subtracting column 6 from column 3. The regression method was not used as the basis for obtaining column 6 (final estimate) for these counties because of unreasonable results in some counties, that is, a negative number of trees in some counties or mail returns of 1 or 2 reports representing extremely large orchards.

duction figures by district were not available; a per orchard estimate was used with the following adjustment for duplication: (1) For each county the number of orchards duplicated in the lists was estimated by dividing the number of sour-cherry trees duplicated by the average number of sour-cherry trees per orchard indicated by the regression method; (2) the number of duplicated sweet-cherry trees was obtained by multiplying the number of orchards found in (1) by the average number of sweet-cherry trees per orchard in the respective counties.

After county estimates were obtained for sours and sweets by the methods outlined above, comparisons were made with 1940 and 1945 census data. Census data by counties seemed to agree reasonably well with the county estimates of total tree numbers, hence the county and district estimates were left unchanged to minimize the use of judgment in deriving final estimates. Judgment estimates were made in only a few counties, but more of such estimates would have been necessary if it had not been for the cooperation of the processors visited, and the availability of inspection data. A summary by districts of the statistics entering into the final estimates is shown in table 2.

Sampling Errors

Sampling errors for the age groups were computed for several counties. At the 5-percent level, or two standard errors, these sampling errors had a range for individual counties from about 10 to 100 percent and averaged about 40 percent; for crop-reporting districts the range was from 10 to 80 percent and averaged about 30 percent; and for the State as a whole the range was from 10 to 30 percent and averaged about 20 percent. For the break-down into only bearing trees and non-bearing trees the sampling errors averaged about 30 percent by counties, 20 percent by districts, and 10 percent for the State.

Remarks

Experience in this survey leads to suggestions that may benefit others who may contemplate similar surveys. They apply to list sampling of the type undertaken in this survey, though they may be of value in other situations. The list of growers was supplied by a State agency which had compiled it from several sources. After it was submitted to county agents, district horticultural agents, and others for review, a practically complete list was obtained. Nevertheless, some omissions and duplications were still present. A combined pre-test of the mailing list and the schedule might be made about as follows:

(1) Select area segments in several counties to test the mailing list for completeness, duplications, and qualifying criteria.

(2) Take a schedule at every 10th farm (or every 5th farm).

Such a scheme would seem desirable if estimates for counties or districts are to be made. Other members of the operator's family or a neighbor could probably give the general infor-

mation required. In some States this kind of checking could be done rapidly by telephone from the county agent's office where township maps are generally available; only those farms from which a schedule was to be obtained would have to be visited personally.

Interviewing alternates at the same time the sample farms were visited, before it was known whether they would be needed, saved many miles of back-tracking, and speeded the interviewing.

The sampling of fruit orchards has shown that orchards vary considerably in size and are frequently fairly scattered. To reduce the costs of travel without sacrificing accuracy the use of the post office or a "postal unit" appears to be desirable. In this survey, after the counties were selected with probability proportional to the number of non-respondents, the interviews were clustered into groups of five farms each by post office or pseudo-post office addresses. Upon selecting names at random for each county, the next four names, following alphabetically in the same post office list, were used to make up clusters of farms. No effects of this clustering could be detected in the mean squares.

The sampling errors of estimates for even very broad categories are rather large for crop-reporting districts and counties. For State totals, the break-down into a large number of age groups as in this survey is likewise subject to rather large sampling errors. We apparently must conclude that if we seek a break-down by a number of varieties or age groups, the results at even the State level will be subject to fairly large sampling errors for the per orchard expansion—say as large as 20 percent. But for broad categories, as total trees, both bearing and non-bearing, the sampling errors for a per orchard expansion can probably be kept within acceptable limits, that is, a 10-percent error at the 5-percent level. If acceptable district estimates are to be obtained, a control factor appears to be necessary. In this survey, the production of sour cherries proved to be a fairly efficient control factor.

Mimeographed indexes for volumes 1 and 2 are now available upon request.