AN ABSTRACT OF THE THESIS OF

Charles Karoly Huszar for the M.S.
(Name of student) (Degree)

in Statistics presented on March 28, 1967
(Major)

Title: THE USE OF OBJECTIVE SAMPLING PROCEDURES IN ESTIMATING OREGON FILBERT PRODUCTION

Abstract approved: Redacted for Privacy
Lyle D. Calvin

In an attempt to improve the mid-season forecast of filbert production, objective measure techniques were started in 1955 on an experimental basis. In the following years an adjusted ratio estimate was used to forecast the filbert production. The purpose of this study was to review the different assumptions and consider different estimating models.

Following a survey of the entire producing area in the Pacific Northwest, a probability sample of trees from 350 orchards was selected. Counts of nuts and defects and nut weights were taken from the sample trees in July of each year to provide data for a forecast of the fall harvest. A subsample of nuts was sized and the size distribution of the nuts examined. A sample of harvested nuts was taken to determine the average nut weights and the defects for each size class.

Results showed that little precision can be gained by geographical stratification. Direct expansion estimates are more precise and yield smaller deviations from the actual production than ratio estimates. The estimates met the objective of a sampling error less than five percent of the production estimate. Larger than expected deviation of the estimates from the actual production can be accounted for by the failure to forecast the changes between survey and harvest time. Examination of size distribution indicates that the estimated percent defects cannot be used alone for forecast the losses after sampling time, because a portion of the good nuts will not be harvested or will be lost during the handling procedures if they are too small to be classified as commercial. Other non-sampling errors are also discussed, but these errors can be minimized by better supervision of the sampling procedures.

It is believed that future investigations should be focused on understanding changes which take place after a sampling time to improve the production estimates. There is little need for further improvement of the precision of the estimates, because the estimates met the objective of a sampling error less than five percent of the production estimate.

THE USE OF OBJECTIVE SAMPLING PROCEDURES IN ESTIMATING OREGON FILBERT PRODUCTION

by

CHARLES KAROLY HUSZAR

A THESIS

submitted to

OREGON STATE UNIVERSITY

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

June 1967

APPROVED:

Redacted for Privacy

Professor of Statistics

In charge of major

Redacted for Privacy

Chairman of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented March 28, 1967

Typed by Joanne Wenstrom for Charles Karoly Huszar

ACKNOWLEDGEMENTS

The author gratefully acknowledges the guidance and suggestions of his major professor, Dr. Lyle D. Calvin. The author also wishes to express his appreciation to the Oregon Filbert Commission, the Filbert Control Board, the Oregon State University Agricultural Experiment Station, and the USDA Agricultural Marketing Service for the research funds which made this study possible and to Mr. Paul Bascom for the organization and supervision of the field work.

TABLE OF CONTENTS

LIST OF TABLES

THE USE OF OBJECTIVE SAMPLING PROCEDURES IN ESTIMATING OREGON FILBERT PRODUCTION

I. INTRODUCTION

In 1955 a research project was initiated to study the use of objective measurements of such filbert crop characteristics as numbers of nuts on the trees, numbers of blanks, nut size and weights, all of which can be observed early in the season, to provide an accurate forecast of the tonnage of filbert nuts moving into trade channels.

The commercial filbert producing area in the Pacific Northwest is limited to a drainage basin in Oregon and Washington called the Willamette-Puget trough. Trees seem to be well adapted to the weather and soil condition of this area. A big acreage increase occurred in the early 1940's, a time when competition was not so keen as now, the total crop was not large, and costs were fairly low. Handsome profits stimulated growers to increase acreage rapidly. In many cases small orchards were planted as part-time operations. On larger farms the crop also provided a means of diversification where land was not well suited to other crops.

Expansion of the industry brought many problems. Growers found that the trees needed special care to maintain high production and quality. Attention needed to be given to pollination problems and spraying became necessary as insects became more numerous. As part of postwar attempts to aid certain European and Mediterranean countries, United States imports of shelled filberts were increased; quota limits were raised and duties were lowered. Year to year fluctuations in the size of the domestic crop made it more and more difficult to meet competition of imports, and filbert production for many small growers became unprofitable. Many plantings were neglected and even removed.

In 1949 the Filbert Control Board was formed, by authority of the Federal Marketing Act, to help growers obtain fair prices without government price supports and to bring some degree of stability to the industry. The function of the Board is to study and recommend the amount of in-shell nuts that can be offered for sale by the trade in any one year, depending upon economic conditions and the domestic supply available. That part of production not going into the domestic in-shell market can be shelled and sold as meats, or be exported, but usually at a lower unit price. According to law, the Secretary of Agriculture must set the percentage of total production to be shelled or exported by late August. He is guided by the recommendation and advice of a board consisting of growers and handlers elected by industry members.

It is readily seen that a considerable amount of responsibility is placed upon one small group of men. Whether too many nuts, or not enough, are held out of the in-shell market, depends almost entirely on the accuracy of the crop forecast. The price level for both domestic and foreign nuts is also usually set early in the season, because the fall and winter holiday markets are the primary selling periods.

A review of official August forecasts made by the Agricultural Marketing Service since 1940 shows the average error to be 12.1 percent. The range is from a low of 2.6 percent in 1947 to 52.7 percent in 1953. During the 17-year period, the error exceeded 10 percent in eight years (13).

In 1955 a cooperative study to improve the forecast of production was started by the Filbert Control Board, the Oregon Filbert Commission, the Oregon Agricultural Experiment Station, and the Crop Reporting Service of the United States Department of Agriculture. The study was conducted with funds from the industry groups and the Experiment Station, matched with Research and Marketing funds from the United States Department of Agriculture (12,13).

The study was set up to attempt to forecast filbert production by measuring year-to-year changes. Observations were made on the count of nuts, size and weight of nuts, and defects such as blanks and worm damages. A ratio estimate was used to forecast production. The general procedure was decided upon because of budget limitation and the apparent success of similar work being conducted in California on peaches and pears and by other investigators in

England and Florida (10,17).

Data obtained in 1955 and 1956 were used to make experimental forecasts of the 1956 crop, and a report was presented to the Filbert Control Board in which four forecasts were presented (15). These were the estimates using nut count ratio only, nut count and dry weight ratio, nut count and dry weight ratio adjusted for change in acreage, and a direct expansion forecast. In the following years more and more attention was paid to an adjusted ratio estimate in which the ratio of nut counts was adjusted for two factors (1). The first is the weight factor, which is basically the ratio of the average dry weight of nuts at a corresponding growth stage. The other factor is the defect factor which corrects the nut count for the estimated percentage of blanks included in the sample.

The following table shows the forecast of production, the final USDA estimated production and the percent deviation of the forecast from the final estimate for the years from 1957 to 1963 (1).

Table I.. Adjusted ratio forecast and final USDA estimate.

Year	Forecast (tons)	Final (tons)	Deviation (%)
1957	13,309	12,510	+ 6.4
1958	7,249	7,540	- 3.8
1959	9,732	10,100	- 3.6
1960	8,165	8,950	- 8.8
1961	10,526	11,760	-10.4
1962	7,848	7,780	+ 0.9
1963	8,200	6,940	+18.2

It should be noted that the high error in the 1963 estimate is due to the overestimation of the effect of a severe October 12, 1962 windstorm. If the allowance made for nuts unharvested in 1962 is excluded, the deviation drops from 18.2 percent to 6.9 percent.

Since data from seven years are now available, it is possible to look at the response of estimating procedures other than those previously used. The purpose of this study is to review the different estimating models. Much of the work includes the computation of the estimates and their variances.

II. BIOLOGICAL DESCRIPTION

Before explaining the selection of the samples as background information relating to the observations made on the sample trees, it may be well to describe the filbert tree, the nut itself, and their growth habits.

The filbert is a small round nut, averaging about 20 mm in diameter, with a relatively hard shell. The flavor of the meat is not strong but has a rather mild and pleasant taste. Commercial varieties grown in this country are all of European origin. Native varieties are commonly called hazel nuts and grow as thickets and brush in many parts of the country.

Trees are usually planted in a uniform pattern, with rows about 20-22 feet apart. By nature the filbert tree is shrublike. If not trained it would grow into an ever-expanding dense aggregate of individual shoots emanating from the ground. For reasons of good cultural practice and a better yield potential, growers prune the plants and train them into the shape of trees with a main trunk and about five main limbs or scaffolds. Continual pruning of shoots that grow up around the trunk is necessary to keep the trees in this condition.

The sex habit of filberts presents problems not ordinarily found in deciduous fruits and nuts. The trees bear separate pistillate and staminate flowers. The pollination season usually begins in December and continues through March. For the main variety (Barcelona), the female flowers are not ready for pollen until the pollen-bearing flowers have passed maturity. Therefore, other varieties are interplanted with Barcelona to provide pollen during the flowering season.

The ovary usually begins to develop several months after pollination and fertilization. When leaf buds swell and begin to form leaves, the female flower also swells and a small shoot develops with some leaves and a small cluster at the tip. The nuts develop in these clusters, with from one to ten per cluster. For most varieties each nut in the cluster is encased in a heavy protective leaf-like husk that grows along with the nut.

The nut usually reaches full size in late July. Then the outer shell starts to harden at the tip and basal end, with hardness spreading from both ends toward the middle. The embryo kernel starts to develop at the blossom end about the time the shell is fully hard. It is connected by a tube (placenta) that extends from the micropyle end of the nut through the filler or packing material inside the shell. Nuts that will be blanks can be identified at shell hardening time by looking at the placenta, which starts to turn brown in such cases.

From this stage on, the meat of the nut develops. During late autumn the mature nuts fall from the husk and are harvested, picked off the ground.

III. SAMPLING

A mail-survey of growers, with follow-up of non-respondents, provided an up-to-date list of all growers that sold nuts through co-operative and independent handlers in either 1953 or 1954. Each planting or block of trees was identified by location, age of trees, variety, and number of trees in each age group. From this listing, arranged by location, age of trees, and size of plantings, a systematic sample of 300 blocks was selected by taking a block for every kth tree after a random start. This provided a stratified sample of blocks with probabilities of selection proportional to numbers of trees. After locating the sample blocks, the orchards were visited and a sketch of each block was drawn to scale, including the number of rows and spaces. A random tree in the orchard was then selected. The location of this sample tree was plotted so samplers could return to the same tree every year. Three trees were sampled, the first at the random position, and two adjacent trees to the east, or to the north if no trees were to the east.

In 1957, an additional 100 sample blocks were selected and 50 blocks were eliminated. Thus, the new sample size became 350 blocks. From this year on, 50 new blocks were selected and 50 blocks were dropped each year. This latter modification was necessary to include new plantings. The increase of the sample size from 300 blocks to 350 blocks was necessary to have 300 matched blocks for the ratio estimates in two consecutive years.

In 1958 and 1963, special tree and acreage surveys were conducted by the Oregon Crop Reporting Service (9) to bring the universe up to date.

Before the field work started, several orchards were visited to determine the "reference date." Rather than using a calendar date, growth stages of the filbert can be used to determine the reference date. The reference date is the calendar date on which 50 percent of the shells of filbert nuts are hard. Prior to 1959, the date on which 90 percent of shells were hard was used.

As soon as the reference date was determined, the field work started. Each sample tree was subdivided into five, approximately equal, parts that could be identified by counting the main branches from the trunk, making allowances, where necessary, for differences in the sizes of those branches. One of those parts was selected at random and all clusters on that part were counted. Every 15th cluster counted was picked. The total cluster count on the sample branches were recorded on the tree card. The clusters picked were also recorded by the number of nuts they contained. These clusters, with the tree card, were packed in plastic bags and sent to Corvallis for further examination.

The laboratory work started upon the arrival of the first samples from the field. All nuts from each tree were weighed as they arrived (green, with husks and bag), then were counted and cracked to determine the percentage of defects. The nuts from the sample bag containing the clusters from the randomly selected tree of each block were peeled, weighed, and sized, then cracked and all the pieces dried for 48 hours at 58 degrees centigrade to obtain the dry weight per sample.

Table II summarizes the average values calculated from the survey data.

Table II. Average values calculated from the survey data.

Year	Clusters per block	Nuts per cluster	Per-cent defect	Dry weight per nut	Refer-ence date*	Average sampling date*
1957	642	2.16	11.0	1.72	10	30.6
1958	471	2.23	20.3	1.65	7	22.2
1959	805	2.00	16.1	1.12	22	29.6
1960	488	2.33	14.0	1.35	13	28.9
1961	730	1.88	19.0	1.32	10.5	26.3
1962	513	1.89	17.9	1.07	19	31.4
1963	412	2.15	19.0	1.22	14	27.4

*Dated from July 1 as day 1.

IV. METHODS OF ESTIMATION

In this study various estimators were employed for the direct expansion, ratio and regression estimates. In these formulas the symbols have the following meanings:

N = total number of trees in the universe
s.u. = sampling unit, group of three trees (described below)
n = number of s.u.
CC = clusters counted per s.u.
CP = clusters picked per s.u.
NC = field count of nuts per s.u.
NP = laboratory count of nuts per s.u.
ND = number of defective nuts per s.u.
WT = average dry weight per nut per s.u.
SD = sampling date of the s.u.
RD = 15 days plus the reference date

The total number of filbert trees were obtained from the report of the Oregon Crop Reporting Service (9). Each sampling unit consisted of a group of three filbert trees. Clusters on one-fifth of each tree were counted, but only every 15th cluster was picked. The field workers classified and recorded the picked clusters by the number of nuts they contained. This permitted the calculation of the number of nuts per cluster; and by multiplying by the cluster count, also recorded by the field workers, the estimate of the average number of nuts per tree can be obtained. Another estimate can be calculated from the laboratory nut count because the clusters picked were separated in the laboratory and an accurate nut count was obtained for each sampling unit.

Defects include both blanks and worm damages. At sampling time the number of defective nuts can be determined relatively well, but defects caused later by worms can alter the production and cannot be included in the estimate. There is also a varying portion of the defects that are not harvested or are blown out before delivery. Ignoring defects after the sampling period is equivalent to assuming that these later defects are proportionately the same in all years. Since it is difficult to get exact figures about the amount of defects in the final production figure, 8.3 percent "allowable defects" were assumed whenever correction was made for defects. This figure was obtained from samples taken from orchard run deliveries (14).

For direct expansion estimates it is necessary to assume an average nut weight to obtain production figures in tons. There are two cases which were considered.

Case A: Correction was made for defects

Case B: Correction was ignored

Case A: When correction is being made for defects, it is assumed that a proportion of the nut count is excluded from the estimate. This is equivalent to assuming that the corrected nut count contains only the allowable defects. After examining data obtained from samples taken from orchard deliveries (14) and from packers (15), 3.0 grams per nut average weight was assumed.

Case B: When correction for defects is ignored, it is assumed that these defects are proportionately the same in all years. It was assumed that ten percent of the nuts will not be harvested or will be blown out before delivery. This is equivalent to assuming that all nuts will be delivered having ten percent lower weight per nut, thus giving 2.7 grams per nut.

In the case of ratio estimates, when correction was made for nut weight differences, it was assumed that the ratio of dry weights 15 days after the reference date is the same as the ratio of the average weight of the harvested nuts. A linear relationship was assumed for the daily increase in dry weight during the sampling period, and was estimated as 0.030 grams per day.

Direct Expansion Estimates

Three forecasting equations to estimate the total production were considered. For each of these models two estimates are presented; one where stratification by area was accounted for in the analysis, and the other where stratification was ignored. This is possible because the number of trees in each area is known.

In the first model the laboratory nut count is used to estimate the average number of nuts per tree. It is assumed that one-fifth of each of three trees were included in the sample and that one-fifteenth of all clusters were picked. An average nut weight is assumed and any change in defects from year to year is ignored.

$$YD(1) = K \frac{\Sigma NP}{n} \tag{1.1}$$

where $K = (15)(5/3)(N)(2.7)$

The second forecasting equation uses nut count estimated by multiplying the cluster count by the average number of nuts per cluster. In this equation the number of nuts in the sampling unit is estimated with the help of the additional information concerning cluster count and nuts per cluster ratio. It is not necessary to assume that every 15th cluster was picked from the sample trees, only that clusters were picked from all trees with equal probability.

$$YD(2) = K \frac{\Sigma CC}{n} \frac{\Sigma NC}{\Sigma CP} \tag{1.2}$$

where $K = (5/3)(N)(2.7)$

In the third model the nut count is corrected for estimated defects. It is assumed that the estimate will contain 8.3 percent allowable defects.

$$YD(3) = K\frac{\Sigma CC}{n}\,\frac{\Sigma NC}{\Sigma CP}\left[1 - \left(\frac{\Sigma ND}{\Sigma NP} - 0.083\right)\right]$$

$$= K\frac{\Sigma CC}{n}\,\frac{\Sigma NC}{\Sigma CP}\left(1.083 - \frac{\Sigma ND}{\Sigma NP}\right) \qquad (1.3)$$

where $K = (5/3)(N)(3.0)$

Ratio Estimates

For ratio estimates of the production, the five estimated ratios are multiplied by the previous year's USDA final production estimates (X), which is based upon delivery records that account for all but a negligible part of the universe. It was impossible to take full advantage of the stratification because the production by area was not available. In the following formulas the subscripts y and x refer to the present and previous years, respectively.

The following five equations were used to estimate production:

$$YR(1) = \frac{\Sigma NP_y}{\Sigma NP_x}\,X \qquad (2.1)$$

$$YR(2) = \frac{\sum CC_y \dfrac{\Sigma NC_y}{\Sigma CP_y}}{\sum CC_x \dfrac{\Sigma NC_x}{\Sigma CP_x}}\,X \qquad (2.2)$$

$$YR(3) = \frac{\sum CC_y \dfrac{NC_y}{CP_y}}{\sum CC_x \dfrac{NC_x}{CP_x}}\,X \qquad (2.3)$$

$$YR(4) = \frac{\sum CC_y \frac{NC_y}{CP_y} \left(1.083 - \frac{ND_y}{NP_y}\right)}{\sum CC_x \frac{NC_x}{CP_x} \left(1.083 - \frac{ND_x}{NP_x}\right)} X \qquad (2.4)$$

$$YR(5) = \frac{\sum CC_y \frac{NC_y}{CP_y} \left(1.083 - \frac{ND_y}{NP_y}\right) \left[W_y + .03(RD_y - SD_y)\right]}{\sum CC_x \frac{NC_x}{CP_x} \left(1.083 - \frac{ND_x}{NP_x}\right) \left[W_x + .03(RD_x - SD_x)\right]} X \qquad (2.5)$$

In equations (2. 1), (2. 2) and (2. 3) any change in defects and in the average nut weight from year to year is ignored. In equation (2. 4) it is assumed that the average nut weight was the same in both years, while in equation (2. 5) the assumption is that the ratio of the dry weights at sampling time is the same as at harvest time. In both equations (2.4) and (2.5) the estimated defects were corrected for allowable defects.

Equation (2. 1) uses the total laboratory nut count and no assumption about the cluster distribution is needed.

In equation (2. 2) the total nut count for the sample is the product of the total cluster count and the nut per cluster ratio of the sample, while in equations (2. 3) through (2. 5) a nut count is obtained for each sampling unit as the product of the cluster count and nut per cluster ratio.

Regression Estimates

It is to be expected that the regression estimate would be more precise than the ratio estimate. Although not much gain in precision is expected, this estimate was mainly included because it is expected to give a better estimate of the population total if the regression line does not pass through the origin, as is assumed in the case of the ratio estimate. To obtain an estimate of the total production the linear regression estimate of the average nut count was multiplied by the same expansion factor as for the direct expansion estimate.

$$YLR = K(A + Bx) = KA + BKx$$

since $Kx = X$

then $YLR = KA + BX$ (3.1)

where $K = (15)(5/3)(N)(2.7)$

X = previous year's actual production

A = least squares estimate of the intercept

B = least squares estimate of the slope of the regression line

Variance Estimates

Procedures to calculate the variances for the functions of two or more random variables are given by Deming (5, page 393). The variance of a forecasting equation was expressed as a linear function of the variances and covariances of its components, and the variance estimate was obtained by substituting the proper sample estimates of the variances and covariances.

Replicated subsampling (4 ,5 ,11) was employed for estimating the variances of the more complex forecasting equations (2. 3),(2.4) and (2. 5). The entire sample was divided into ten subsamples , then the estimated nut counts for the present and past years were calculated for each sampling unit. The estimate of the variance of the ratio was obtained from the variance formula given by Deming (5, page 199).

V. RESULTS

Tables III through IX display the different types of estimates for each calendar year from 1957 to 1963. The identification of the estimates refers to the forecasting equation of the previous chapter. For direct expansion estimates, two estimates are presented: (a) represents when stratification by area was considered, and (b) when stratification was ignored in the calculation of the estimate of production and its sampling error.

Data from 1956 were not complete; therefore, for 1957, only direct expansion estimates were calculated, and they are presented in Table III. These estimates are based on 331 sampling units.

Table III. Production estimates, 1957.

Type of estimate	Estimated production (Tons)	Deviation from final (%)	Sampling error (Tons)	Sampling error (%)
1.1(a)	12,913	+ 3.2	390	3.0
(b)	13,070	+ 4.5	416	3.2
1.2(a)	11,086	-11.4	361	3.3
(b)	11,161	-10.8	381	3.4
1.3(a)	12,155	- 2.8	396	3.3
(b)	12,245	- 2.1	418	3.4
Final	12,510	-----	---	---

Table IV displays the results for 1958. Three hundred forty-nine sampling units were used to calculate the expansion estimates, and data from 287 blocks were matched with the 1957 data to be used to calculate the ratio estimates.

Table IV. Production estimates, 1958.

Type of estimate	Estimated production (Tons)	Deviation from final (%)	Sampling error (Tons)	(%)
1.1(a)	8,048	+ 6.7	283	3.5
(b)	8,069	+ 7.0	312	3.9
1.2(a)	8,017	+ 6.3	282	3.5
(b)	8,037	+ 6.6	312	3.9
1.3(a)	7,948	+ 5.4	287	3.6
(b)	7,973	+ 5.7	320	4.0
2.1	7,765	+ 3.0	364	4.7
2.2	9,268	+22.9	465	5.0
2.3	9,266	+22.9	368	4.0
2.4	8,402	+11.4	348	4.1
2.5	11,429	+51.6	543	4.7
3.1	7,777	+ 3.1	317	4.1
Final	7,540	-----	---	---

In 1959, 346 blocks were used for direct expansion estimates from which 294 were paired with the previous year. Table V shows the results.

Table V. Production estimates, 1959.

Type of estimate	Estimated production (Tons)	Deviation from final (%)	Sampling error (Tons)	(%)
1. 1(a)	11,960	+18.4	369	3. 1
(b)	11,912	+17. 9	368	3. 1
1. 2(a)	12,109	+19. 9	373	3. 1
(b)	12,054	+19. 3	371	3. 1
1. 3(a)	12,610	+24. 8	394	3. 1
(b)	12,540	+24. 2	393	3. 1
2. 1	11,290	+11. 8	460	4. 1
2. 2	11,471	+13. 6	469	4. 1
2. 3	11,451	+13. 4	434	3. 8
2. 4	11,975	+18. 6	468	3. 9
2. 5	9,285	- 8. 1	426	4. 6
3. 1	11,905	+17. 9	474	5. 0
Final	10,100	-----	---	---

Of the 345 sampling units 290 were matched with blocks in 1960. The results are given in Table VI.

Table VI. Production estimates, 1960.

Type of estimate	Estimated production (Tons)	Deviation from final (%)	Sampling error	
			(Tons)	(%)
1.1(a)	8,853	- 1.0	276	3.1
(b)	8,940	- 0.1	308	3.4
1.2(a)	8,640	- 3.5	293	3.4
(b)	8,733	- 2.4	327	3.7
1.3(a)	9,184	+ 2.6	332	3.6
(b)	9,290	+ 3.8	367	3.9
2.1	7,806	-12.8	300	3.8
2.2	7,558	-15.6	306	4.1
2.3	7,565	-15.5	288	3.8
2.4	7,763	-13.3	302	3.9
2.5	7,490	-16.3	253	3.4
3.1	8,488	- 5.2	312	3.5
Final	8,950	-----	---	---

In 1961, 342 blocks were considered in the direct expansion estimates and 296 of those were sampled in the previous years.

Table VII. Production estimates, 1961.

Type of estimate	Estimated production (Tons)	Deviation from final (%)	Sampling error (Tons)	(%)
1. 1(a)	11,127	- 5.4	320	2. 9
(b)	11,044	- 6. 0	319	2. 9
1. 2(a)	10,595	- 9. 9	305	2. 9
(b)	10,507	-10. 6	305	2. 9
1. 3(a)	10,677	- 9. 2	320	3. 0
(b)	10,580	-10.0	322	3. 0
2. 1	11,173	- 5.0	485	4. 3
2. 2	10,847	- 7.8	501	4.6
2. 3	10,849	- 7.8	479	4.4
2.4	10,255	-12.8	474	4.6
2. 5	10,096	-14.2	403	4.0
3. 1	11,242	- 4.4	346	3. 1
Final	11,760	-----	---	---

Only 330 blocks gave useful information in the 1962 survey; and from these, 286 were matched for ratio estimate. The estimates are displayed in Table VIII.

Table VIII. Production estimates, 1962.

Type of estimate	Estimated production (Tons)	Deviation from final (%)	Sampling error (Tons)	(%)
1. 1(a)	7,812	+ 0.4	243	3. 1
(b)	7,718	- 0. 8	263	3. 4
1. 2(a)	7,445	- 4. 3	254	3. 4
(b)	7,379	- 5. 2	278	3. 8
1. 3(a)	7,598	- 2. 3	302	4. 0
(b)	7,521	- 3. 3	334	4. 4
2. 1	8,474	+ 8. 9	326	3. 8
2. 2	8,521	+ 9. 5	358	4. 2
2. 3	8,266	+ 6. 2	306	3. 7
2. 4	8,330	+ 7. 1	329	4. 0
2. 5	7,456	- 4. 2	290	3. 9
3. 1	8,118	+ 4. 3	271	3. 3
Final	7,780	-----	---	---

In 1963, the sample size was increased and the expansion estimates were based on 380 sampling units and the ratio estimates on 316 units. The following table shows the estimates for this year.

Table IX. Production estimates, 1963.

Type of estimate	Estimated production (Tons)	Deviation from final (%)	Sampling error (Tons)	(%)
1. 1(a)	7,153	+ 3.1	221	3.1
(b)	7,257	+ 4.6	274	3.8
1. 2(a)	6,482	- 6.6	215	3.3
(b)	6,577	- 5.2	258	3.9
1. 3(a)	6,531	- 5.9	224	3.4
(b)	6,623	- 4.6	269	4.0
2. 1	7,747	+11.6	433	5.6
2. 2	7,268	+ 4.7	430	5.9
2. 3	7,406	+ 6.7	377	5.1
2. 4	7,262	+ 4.6	384	5.3
2. 5	7,377	+ 6.3	438	5.9
3. 1	7,339	+ 5.7	303	4.1
Final	6,940	-----	---	---

The following two tables compare the percent deviations and the sampling errors in percent of the estimates from 1958 to 1963.

Table X. Percent deviation of the estimates from the actual production.

Type of estimate	1958	1959	1960	1961	1962	1963
1.1(a)	+ 6.7	+18.4	- 1.0	- 5.4	+ 0.4	+ 3.1
(b)	+ 7.0	+17.9	- 0.1	- 6.0	- 0.8	+ 4.6
1.2(a)	+ 6.3	+19.9	- 3.5	- 9.9	- 4.3	- 6.6
(b)	+ 6.6	+19.3	- 2.4	-10.6	- 5.2	- 5.2
1.3(a)	+ 5.4	+24.8	+ 2.6	- 9.2	- 2.3	- 5.9
(b)	+ 5.7	+24.2	+ 3.8	-10.0	- 3.3	- 4.6
2.1	+ 3.0	+11.8	-12.8	- 5.0	+ 8.9	+11.6
2.2	+22.9	+13.6	-15.6	- 7.8	+ 9.5	+ 4.7
2.3	+22.9	+13.4	-15.5	- 7.8	+ 6.2	+ 6.7
2.4	+11.4	+18.6	-13.3	-12.8	+ 7.1	+ 4.6
2.5	+51.6	- 8.1	-16.3	-14.2	- 4.2	+ 6.3
3.1	+ 3.1	+17.9	- 5.2	- 4.4	+ 4.3	+ 5.7

Table XI. Percent sampling error of the estimates.

Type of estimate	1958	1959	1960	1961	1962	1963
1.1(a)	3.5	3.1	3.1	2.9	3.1	3.1
(b)	3.9	3.1	3.4	2.9	3.4	3.8
1.2(a)	3.5	3.1	3.4	2.9	3.4	3.3
(b)	3.9	3.1	3.4	2.9	3.8	3.9
1.3(a)	3.6	3.1	3.6	3.0	4.0	3.4
(b)	4.0	3.1	3.9	3.0	4.4	4.0
2.1	4.7	4.1	3.8	4.3	3.8	5.6
2.2	5.0	4.1	4.1	4.6	4.2	5.9
2.3	4.0	3.8	3.8	4.4	3.7	5.1
2.4	4.1	3.9	3.9	4.6	4.0	5.3
2.5	4.7	4.6	3.4	4.0	3.9	5.9
3.1	4.1	5.0	3.5	3.1	3.3	4.1

Table XII is a summary of Tables X and XI. It displays for each type of estimate the average value of the percent deviations and the corresponding pooled sampling error. It also shows the average value of absolute percent deviation and the corresponding sampling error.

Table XII. Summary of the estimates.

Type of estimate	Average error		Average absolute error	
	Deviation	S. error	Deviation	S. error
1. 1(a)	3. 7	3. 1	5. 8	2. 5
(b)	3. 8	3. 4	6. 1	2. 7
1. 2(a)	0. 1	3. 3	8. 6	2. 6
(b)	0. 4	3. 5	8. 2	2. 8
1. 3(a)	0. 9	3. 5	8. 4	2. 8
(b)	2. 6	3. 8	8. 6	3. 0
2. 1	2. 9	4. 4	8. 8	3. 5
2. 2	4. 5	4. 7	12. 4	3. 8
2. 3	4. 4	4. 2	12. 1	3. 4
2. 4	2. 6	4. 3	11. 3	3. 5
2. 5	2. 5	4. 5	16. 8	3. 6
3. 1	3. 6	3. 9	6. 8	3. 1

VI. DISCUSSION OF THE RESULTS

The problem of making accurate annual estimates of the total filbert production, using data obtained in the July survey, is not only a sampling problem. Biological or weather factors can affect the final or actual production after the survey period, and may cause, indirectly, an "error" in the estimate. This may have happened in 1959, because all but one estimating equation considerably overestimates the production, as it can be seen in Table X. In 1960, on the other hand, while the direct expansion estimates are very close to the actual production, the ratio estimates considerably underestimate the production value. If this reasoning is valid, the estimates should have been compared to a "corrected actual production," which would account for irregularities between sampling and harvest time; and this figure should be used for the next year's ratio estimates. On the other hand, when irregularities are observed after the July forecast is made, it should be brought to the attention of the filbert industry, or a revised estimate be made.

Of course, irregularities sometimes can be misleading. For instance, it was believed after the October windstorm in 1962, which caused a considerable amount of damage in the orchards, that a portion of the crop was not harvested, and adjustment was made for this amount when the 1963 estimate was reported as it is shown in Table I.

It turned out that the adjustment was unwarranted. A special tree and acreage survey (9) of the orchards in mid-1963 showed that the total immediate loss amounted to 10.8 percent of the number of trees standing before the storm, but 5.3 percent of the total was reset; and, thus, only 5.5 percent was removed. It can be argued that this loss was even less because the "natural thinning" of the windstorm increased quality of the orchard directly, or indirectly by forcing the owner to spend more time in the orchard.

The above discussion explains some of the non-statistical factors, which might account for larger than expected deviations in several years. It also indicates that further detailed studies are necessary to understand completely the factors which alter the production after the July survey. It should be emphasized again, that the estimated production is not compared with the true population production because it is not available. The USDA final production estimate was used as a basis of the comparison. This estimate is believed to be very close to the true population value.

Table XII displays the average values of the yearly estimates for the time period from 1958 to 1963 inclusive. The average deviation is the arithmetic mean of the deviation and average absolute deviation or mean deviation is the mean of the absolute values of the yearly deviations. The sampling error of the average deviation is the square root of the average relative variance. The error associated with the average absolute deviation is 0.80 times the average sampling error, which is the expected value of the mean deviation if normality is assumed. The following evaluation of the different types of estimating equations will be based mainly on their precisions, that is, on their sampling errors and only secondary consideration will be made about their deviation from the actual.

Direct Expansion Estimates

Comparing the estimates (a) and (b) for the direct expansion estimates, it is obvious that there is little difference in precision, and the estimates are practically identical. Taking advantage of the stratification by area appears to yield less than ten percent increase in precision. The average yield per tree does not seem to vary much by location.

It was expected that estimate (1.2) would be superior to (1.1), because estimate (1.1) assumes that exactly one-fifteenth of the clusters of the sampling units are picked while estimate (1.2) considers this sampling ratio as variable. The added variation can account for the slightly higher sampling error. Considering the deviations, estimate (1.1) tends to be larger than (1.2); but the larger average absolute deviations of (1.2) indicate less consistency than for estimate (1.1). This failure probably is due to a non-sampling error committed when the field worker classifies the clusters by number of nuts per cluster.

Formula (1. 3) does not seem to verify the expectation that correction for defect improves the estimate. This is probably due to the assumption concerning the allowable defect, which may be correlated with the final production instead of being constant from year to year. A more detailed study is needed to answer this question.

Ratio Estimates

Table XII indicates that the ratio estimates are generally less precise than the direct expansion estimates. There could be two reasons for this. One, the ratio estimates are calculated from a smaller sample, about 300 sampling units, while the direct expansion estimates are based on about 350 sampling units. The other reason might be that the low correlation between the two matching samples in some years does not justify the use of ratio estimates. Cochran (3, page 200) gives conditions under which the ratio estimates are more precise than expansion estimates.

Since the ratio estimates yield about the same sampling error, they will be compared by their average absolute deviation shown in Table XII. Formula (2. 1) has the smallest mean deviation and Formula (2. 5) gives the largest one. They are at the same time the simplest and most complex expression respectively. Formulas (2. 4) and (2. 5) do not seem to verify the validity of the correction attempt made by the defect factor, and the assumption for (2.5), that the corrected nut weight ratio at sampling time is a good estimate of the weight ratio at harvest time, seems to be unwarranted. Equations (2.1), (2.2) and (2.3) are basically the same expressions, and (2.1) probably the best because the nut counts were obtained in the laboratory, while the other two equations may involve recording errors of the field worker.

Regression Estimates

The regression estimate (3.1), as expected, is more precise than the ratio estimate (2.1) but not as precise as the direct expansion estimate (1.1). Both the average and the mean deviation of the regression estimate lie between the corresponding deviations of estimate (1.1) and (2.1), indicating that the regression estimate includes features of both types of estimates.

Preferred Estimate

Estimate (1.1) is the most precise among all investigated estimates. Its average absolute deviation is also the smallest as it can be seen in Table XII. This fact does not mean that more complicated estimating formulas are unwarranted, but it does mean that complete understanding of the behavior of the other factors involved are necessary to arrive at a better estimating equation. Estimate (1.1) is better than (2. 1) because it uses, as an expansion factor, the total number of the filbert trees in the universe. This does not fluctuate from year to year as much as the previous year's production, which is used to multiply the nut ratio by to obtain estimate (2. 1). It seems that less error is introduced when a general assumption about defects and average nut weight is assumed, than when an attempt is made to estimate these population parameters by biased estimates.

Multiple Regression Analysis

Stepwise multiple regression analysis was used in an attempt to determine a functional relationship between the USDA final production estimate and the estimates obtained from the sample. The following variables were used as independent variables: laboratory nut count, percent defect, average dry nut weight at sampling time, and the difference between the average sampling date and the reference date. Survey years 1957 through 1963 were used. It is necessary to emphasize that this analysis assumes the same linear relationships in every production year which seems unlikely. The results showed that 86. 5% of the total variance of the production estimate can be explained by the variation in nut count; 8. 3%, 2.4%, and 1. 6% are due to the average sampling date, average nut weight, and defects respectively, and only 1. 2% is unaccounted for and assumed to be random variation. If this analysis has some meaning, then it suggests that a better control of the sampling period is necessary, that is, the average sampling date should be at a set time after the reference date.

Size Distribution

One defect of the ratio method is that the weight ratio at sampling time does not estimate the harvest weight ratio accurately enough. By assuming that the same size nuts will weigh the same every year, a shift in the size distribution of nuts can account for the change in average weight.

Since 1957, a subsample of the nuts sent to the laboratory has been sized to obtain the expected size distribution of Barcelona nuts. Prior to 1962, a sizing ring was used to classify the nuts into four commercial grades only. Since 1962, a new sizing ring has been used to obtain a better size distribution by classifying the nuts into 16 size classes.

To gain more knowledge about the distribution, the nuts of the 1964 sample, after sizing, were cracked and the defects were recorded by size groups. For comparative reasons, a sample of the 1964 crop was obtained primarily to establish average weights for each size class. Subsamples from each size class were cracked to determine the percentage defects. Table XIII displays the results obtained from the sizing and cracking tests of the July sample, while

Table XIV shows the results of the tests made on the samples of the crop. Since comparisons can be made only for the 1964 production year, more similar experiments are necessary to understand the relation between the July and harvest sample.

Comparing Tables XIII and XIV, it can readily be seen that the small nut size classes A, B, C and D include from 8 to 17 percent of the nuts taken from the survey samples and these size classes are almost entirely missing from the sample taken from the 1964 harvested nuts. These groups contain mostly defective nuts; but, since they are missing from the harvest sample, all these nuts should be considered as defects. On the other hand, defects in the survey and in the harvest sample seem to decrease as the sizes of the nuts are increasing, and does not seem to verify the assumption that change in the size distribution is due to the loss of the defective nuts.

The above discussion indicates that a varying portion of the estimated defects at survey time will be lost at harvest time, and in addition to this loss, most of the good nuts from the small size groups will be lost due to the harvesting and handling procedures.

Table XIII. The size distribution of nut samples at survey time.

Size class	Upper limit (mm)	--- YEARS --- 1962 (%)	1963 (%)	1964 (%)	Defects 1964 (%)
A	12.7	11.2	7.9	5.6	96
B	13.7	1.2	1.0	0.5	85
C	14.6	1.6	1.3	0.7	67
D	15.5	3.0	1.9	1.2	56
E	16.5	4.7	2.8	1.9	49
F	17.2	4.2	3.3	3.1	44
G	18.0	6.5	5.1	5.2	33
H	18.7	9.0	8.6	10.8	26
I	19.5	13.6	11.5	13.2	21
J	20.2	15.4	14.4	18.3	17
K	21.2	18.4	19.7	19.4	16
L	22.2	8.3	13.7	12.2	15
M	23.0	2.2	6.6	5.1	19
N	23.8	0.7	2.1	2.5	21
O	24.5	0.0	0.1	0.2	14
P	----	0.0	0.0	0.1	17

Table XIV. The size distribution of 1964 harvested nuts.

Size class	Size by number (%)	Distribution by weight (%)	Defect by number (%)	Weight per nut (grams)	Adjusted weights (grams)
A	----	----	--	---	0.5
B	----	----	--	---	0.8
C	0.03	0.01	0	1.1	1.0
D	0.20	0.08	32	1.1	1.2
E	0.74	0.36	39	1.4	1.4
F	1.33	0.74	18	1.6	1.6
G	3.97	2.47	23	1.8	1.8
H	8.14	5.92	18	2.1	2.1
I	13.99	11.56	18	2.4	2.4
J	27.50	26.49	13	2.7	2.7
K	28.11	31.52	9	3.2	3.2
L	12.03	15.22	9	3.6	3.6
M	3.05	4.24	10	3.9	3.9
N	0.65	0.98	13	4.3	4.3
O	0.17	0.27	4	4.5	4.5
P	0.09	0.14	0	4.6	4.7

Average weights were calculated for most of the size classes, and adjustments were made graphically to extend to all size classes. The average weight for the 1964 harvest sample was 2.84 grams per nut and the same value was obtained from the survey sample if size classes A through D were ignored. Including these classes, an average weight of 2.59 grams per nut was obtained, which underestimates the harvest average weight.

It seems that the size distribution and the established class nut weight may yield a good estimate of the average harvest weight per nut, but this weight must be used with nut counts which are corrected for losses instead of defects. For example, size classes A through D may be deleted as losses.

Non-Sampling Errors

It may be well to discuss some of the non-sampling errors that exist in the sampling scheme being tested or observed. It is difficult to know all that do play a part, and of those that can be enumerated, just how important they are. Some places that non-sampling errors seem most apparent are the following:

Selection of the Tree for Sampling. The sampler, while instructed to select trees in a predescribed manner, may in a few cases where there is a choice, select the smaller tree or easier one to count. For example, where a randomly selected tree was actually missing, he was to take the next tree to the south. There were one or two cases where the point of entry to the orchard, using the random row and space as selected, would have put the samplers in a fairly difficult position to get to, and consequently a new point of entry was made by the sampler.

Selection of the Part of the Tree to be Sampled. No doubt the most obvious of possible sources of non-sampling errors is the failure of the sampler to select one-fifth of the tree to be sampled. However, this is probably not too important when using the ratio estimate since the exact same tree and part of tree is sampled every year. But if a sampler, for example, consistently selects a smaller part of the tree the direct expansion estimate will be influenced by his error.

Actual Counting of Clusters by the Sampler. Another very obvious possible source of bias is in the cluster count. The sampler has a difficult task to sample the tall trees or the top of medium size trees, thus he may undercount the clusters, or he may correctly count the clusters but pick too few. The former error will give an inaccurate count, while the latter might give the correct cluster count but will yield an inaccurate nut per cluster ratio. In order to give some check on the cluster count, each sampler had at least five trees sampled at different times during the period of sampling on which he counted one tree, then stripped and compared counts. In two consecutive years, the counting error was a 12 percent undercount for the on-tree count. If it can be assumed this represents the average counting error and it was the same in both years, no adjustment is necessary for the ratio estimates, but it would be necessary to adjust the direct expansion estimates.

Nut Count. One type of nut count used in the estimates is the product of the cluster count and the nut per cluster ratio. In addition to the possible errors described in the previous paragraph, the sampler may fail to classify the larger clusters into the correct nut per cluster group, by overlooking small nuts. The other type of nut count used is based on the actual count of nuts in the laboratory where the clusters were separated and counted. This laboratory count was higher than the sampler's count in five out of seven years.

Reference Date. Since the start of the sampling period is determined by the reference date, the nuts in the first trees sampled are less mature than the nuts sampled later. This can cause an error in the detection of defects, thus giving an incorrect estimate of the defects in the sample. If the reference date is determined incorrectly, an error of one day results in an error of three percent in the estimate.

Defects. In the estimating formulas when correction was made for defects, a constant allowable defect percentage was assumed, and the excess in the estimated defects was assumed to be a loss.

The examination of the size distribution does not seem to verify these assumptions, as was previously discussed. This is a possible source of bias in the estimates, even if the estimate of defect percentage at sampling time is correct.

Acreage Change. It is known that the bearing acreage is also changing from year to year. For ratio estimates the acreage change was ignored because no estimate of the year to year change was available. The estimate of the total number of trees in the universe was used for the direct expansion estimates for expansion factor. This number was assumed unchanged from the previous tree and acreage survey; that is, no change was assumed between survey years 1955, 1958 and 1963 because of the lack of any knowledge of the type of change.

VII. SUMMARY

The problem of making accurate annual estimates of the total filbert production has been examined in this study by investigating several different methods of estimating production and their related errors. These estimates generally had about the same sampling error, but the estimates of the total production, due to different biases, gave larger than expected deviations from the actual production.

The more complex estimating equations yielded larger deviations from the actual production than the less complex ones. An attempt was made to estimate loss between sampling and harvest time by assuming that a portion of the defective nuts detected at sampling time would be lost and thus decrease the estimate. This assumption did not seem to be verified because other losses, such as unharvested good nuts, might differ in different production years. Another assumption made for equation 2.5, that the ratio of the average dry weight at sampling time is the same as the ratio of the harvested nut weight, was not warranted.

Some of the non-sampling errors due to the samplers' failure to follow instructions were examined. These errors may yield minor biases, but by more direct supervision, can be eliminated. It is believed that the sampling procedure is satisfactory. Work by others on this project indicates that the estimates may be improved by sampling one-tenth of six trees instead of one-fifth of three trees per sampling unit.

The major source of bias is the estimation of the possible changes between sampling and harvest time, especially the losses due to harvesting and handling procedures. It is believed at this stage that useful information can be gained through the examination of changes in the size distribution after sampling time. From the size distribution the portion of the small nuts can be determined and be classified as losses, instead of using the estimate of defects for this purpose. After these probable losses are determined, and assuming that the same size nuts weigh the same every crop year, the average weight per nut can be calculated.

BIBLIOGRAPHY

1. Bascom, Paul F. and Lyle David Calvin. 1964 Filbert forecasting research project. Portland, Oregon, Oregon Crop Reporting Service, August 1964. 2 p. (Mimeographed Report)

2. Calvin, Lyle David. Discussion: Forecasting yields with objective measurements. Journal of Farm Economics 45:1513-1514. 1963.

3. Cochran, William G. Sampling techniques. 2d ed. New York, Wiley, 1963. 413 p.

4. Deming, William Edwards. On simplification of sampling design through replication with equal probabilities and without stages. Journal of the American Statistical Association 51:24-53. 1956.

5. Deming, William Edwards. Sample design in business research. New York, Wiley, 1950. 517 p.

6. Deming, William Edwards. Some theory of sampling. New York, Wiley, 1950. 602 p.

7. Hansen, Morris H. , William N. Hurwitz and William G. Madow. Sample survey methods and theory. Vol. 1. New York, Wiley, 1953. 638 p.

8. Hendricks , Walter A. Forecasting yields with objective measurements. Journal of Farm Economics 45:1508-1513. 1963.

9. Hile, Robert B. and Paul F. Bascom. 1963 Filbert tree and acreage survey. Portland, Oregon, Oregon Crop Reporting Service, 1963. 3 p. (Mimeographed Report)

10. Jessen, R. J. Determining the fruit count on a tree by randomized branch sampling. Biometrics 11:99-109. 1955.

11. Jones, Howard L. Investigating the properties of a sample mean by employing random subsample means. Journal of the American Statistical Association 51:54-83. 1956.

12. Parr, Robert D. and Lyle David Calvin. A new method of forecasting filbert production. In: 41st Annual Proceedings of Northwest Nut Society, December 1955. Portland, Oregon, 1955. p. 193-196.

13. Parr, Robert D. and Lyle David Calvin. Forecasting filbert production from counts of nut set and weights of nuts. Paper presented at the meeting of the American Statistical Association, Atlantic City, New Jersey, September 10, 1957.

14. Parr, Robert D. and Lyle David Calvin. 1956 Filbert forecasting research project, Preliminary Report. Portland, Oregon, Oregon Crop Reporting Service, August 1956. 6 p. (Mimeographed Report)

15. Parr, Robert D. and Lyle David Calvin. 1957 Filbert forecasting research project. Portland, Oregon, Oregon Crop Reporting Service, August 1957. 4 p. (Mimeographed Report)

16. Sukhatme, Pandurang V. Sampling theory of surveys with applications. Ames, Iowa, The Iowa State College Press, 1954. 491 p.

17. Yates, F. Crop prediction in England. Bulletin of the International Statistical Institute 33: 295-311. 1951.