Report of California Methodological Projects
by Wendell Wilson
January 26, 1972

1. Fruit and Nut Acreage Sampling Project

The fruit and nut acreage sampling project consisted of the construction and evaluation of a specialized area sampling frame for estimating current fruit and nut crop inventories. The sampling units of the frame were basically sections of land with fruit and/or nut acreage. For each sample unit there was an abundance of very detailed supplemental data on fruit and nut acreages in recent years. The availability of a specialized frame with detailed supplemental data suggested an opportunity for sampling efficiencies in estimating detailed and current inventories of tree and vine crops.

Before I go into more detail on this project, allow me to briefly acquaint you with the present system of collecting fruit and nut acreage data in California. The present system is, of course, the one for which this project was designed to seek an alternative.

The present system, the regular fruit and nut acreage program, consists of census type surveys, performed county by county, with a complete cycle covering the entire State about once every five years. As a result, the survey output never provides a current inventory at a given time. Estimates of inventory change in the interim between censuses are provided by the county agricultural commissioners. However, the job they do in reporting new plantings and removals is very inconsistent from county to county. Their effort is funded by the counties so that neither the State Department of Agriculture nor the California SSO has any direct control. Because most of the lag in the fruit and nut inventory is made up of recently planted (non-bearing) trees and vines, it has little effect on estimated production. The lag for many counties is not fully reduced until the census of the county is conducted. This can be zero to five years after planting for an individual county or an average of about two and a half years at the State level.

During fiscal 1971 priority was given to evaluation of the completed portion of the fruit and nut acreage sampling frame, which was constructed during the two preceding years. The frame was constructed by summarizing data from the regular fruit and nut program by sample units specified as part of the frame. In regular land grid areas an A.D.P. program summarized data from the most recent county census by section, township and range numbers. In other areas aerial photographs and descriptions of the road location of blocks of fruit and nuts were used to construct sample units of approximately 640 acres.

The evaluation was limited to single county to allow a more detailed study of the utility of the frame for estimating fruit and nut acreage. San Joaquin County was selected because of its importance in the production of fruit and nuts. The evaluation consisted of three phases.

Phase I. Analysis of the frame based upon the 1967 data from which it was constructed.

Phase II. Simulation of various alternative sampling designs using the 1967 frame to sample data collected by the latest regular fruit acreage enumeration in 1970.

Phase III. Utilization of the frame to test the efficiency and feasibility of a promising sample design by conducting a pilot survey.

Phase I involved frequency distributions of total fruit and nuts, individual crop and variety acreage. The frequency distributions were completed for all sample units, sample units by total fruit and nut acreage strata (four of them), and sample units by total fruit and nut acreage and cherry acreage strata (11 of them). This provided some insight into the form of the total acreage, individual crop acreage and individual variety acreage populations and how they were affected by various types of stratification. Also, an acreage correlation matrix was computed for 10 major crops and total fruit and nuts. The frequency distributions and correlation matrix were useful in indicating efficient types of sample designs and estimators for estimating San Joaquin County fruit and nut acreage.

Variances, standard errors and coefficients of variation were calculated for 19 crop classes for each of 10 different sampling designs. All computations were based on a sample of 62 of the 625 sample units in the county. Stratification by total fruit and nut acreage compared to a simple random design greatly increased the precision of estimated acreage for most crops, but the indicated precision of estimated cherry acreage was unacceptably low. Because San Joaquin County produces the majority of California's cherries, it was necessary to increase the precision of the estimate for the county, so that it would be compatible with the precision needed for a State estimate. This was the basic reason for stratification by cherry acreage. Proportional allocation consisted of the same sampling rate (10%) for each stratum. The individual optimum was the allocation best suited to estimating for each individual crop, but not necessarily suited to estimating other crops. Total fruit and nut optimum estimated the total best, but, of course, that estimate is of doubtful value. The compromise allocation was a compromise of the individual allocations of almonds, English walnuts, table grapes, wine grapes and cherries. It seemed to provide estimates for most more important crops with precision comparable to their individual allocations. The compromise - optimum allocation for the 11 strata appeared to provide the most precise estimate for most major crops.

Phase I involved indications of the precision attained with various sampling designs based upon sampling of a population with a frame containing supplemental information which was completely current. This is a somewhat artificial situation, but does provide a valid means of comparing alternative sampling designs. Phase II (simulation of various alternative sampling designs using the 1967 frame to sample data collected by the latest regular fruit acreage census in 1970) provided indications of how changes over time affect the efficiency of various sampling designs. Phase II, also, provided indications of the efficiency of using various types of systematic sampling. Seven different sampling designs were simulated by selecting five independent samples of 62 sample units for each design. One sample was a systematic sample with the sample units arrayed in sample unit order. Because sample units geographically near to each other are also numerically close, systematic selection forces a degree of geographic dispersion in the sample. The other six designs were variations of sample allocations and methods of sample selection within strata for a stratified sample by total fruit and nut acreage and cherry acreage (11 strata). Three designs involved proportional allocation and three utilized the compromise-optimum allocation developed in Phase I. For each type of allocation, one design involved random selection within strata. The other two designs involved systematic sampling within strata; one with frame ordered by sample unit number and the other with the frame ordered by total fruit and nut acreage. Compromise-optimum allocation appeared to provide the greatest precision for most crops, especially the major crops. It was not clear, however, which type of selection within strata provided the most precise estimates.

Phase III (utilization of the frame to test the efficiency and feasibility of a promising sample design by conducting a pilot survey) was carried out for a multiple-stratified sample (12 strata) with compromise-optimum allocation and systematic selection of sample units arrayed by total fruit and nut acreage. The 11 strata are those previously discussed. The 12th stratum was composed of sample units enumerated as having some fruit and nut acreage in 1970, but none in 1967. This stratum accounts for fruit and nut acreage not on the original frame which needs to be identified prior to sample surveys so that the complete population can be covered. The population sampled included 670 units, and accounts for all fruit and nut acreage in the county in sample units enumerated as having fruit and nut acreage in 1967 and/or 1970. The sample included 67 sample units (a 10 percent sample).

Collection of data for the pilot survey was completed in three weeks during June. Aerial photographs were used for most sample units. Each operation within a sample unit was accounted for by drawing off its boundaries on the photo after its fruit and nut inventory had been reported. This provided a feasible means of insuring that all fruit and nuts were reported in each sample unit. Detailed orchard record data was obtained for each operation. Owner and/or operator's name and address, property location, total acres in the operation, and for each planting; crop, variety, planting year, planting system, planting space, number of trees (or vines) and acres were obtained. Field enumerators had no difficulty in locating sample units and boundaries were easy to determine in most cases.

The pilot survey also provided cost data indications. Indicated average cost per sample unit including salary of enumerators, travel and office costs ($5 per sample unit) were: $24, $33, $33, $35, and $38 for sample units with 1967 fruit and nut acreage of zero, 0.1-100.0, 100.1-250.0, 250.1 to 500.0 and more than 500.0, respectively. This indicates a cost of about $31,000 for a 10 percent sample of fruit and nut areas of the State. This would not include costs of preparing or updating the sampling frame. Periodic enumeration of counties and/or aerial photography coverage might be necessary to maintain a sufficiently current sampling frame.

Although the pilot survey was not designed to provide adequate precision at the county level, five indications of acreage for 21 crops and crop groupings were computed. Also, variety acreages were estimated by direct expansion. The indications were the direct expansion, 1971/67 ratio estimates by crop and by total fruit and nut acreage and 1971/70 ratio estimates by crop and by total fruit and nut acreage. Also, June enumerative segments in San Joaquin Coutny (49) were expanded to provide an independent estimate. The estimates, while reasonably consistent, cannot be compared with know totals as the 1970 simulated estimates were. The estimates for grapes were compared to results of a special grape acreage survey. Totals by types and some by individual varieties were in close agreement. However, as might be expected, for some varieties the acreage estimates were far apart.

One of the main benefits of this project was the development of procedures for selecting an efficient and feasible sampling design utilizing a sampling frame which can be initially constructed from the present system. If funding can be obtained for constructing the frame, it was recommended that conversion to a sample survey estimating program be initiated by regions of the State. Any sample survey cannot feasibly provide as detailed an inventory record as the present program, but will provide current estimates with sufficient detail and reliability for regions of the State and at the State level.

2. Special Grape Acreage Survey

The special grape acreage survey was another alternative or supplement to the regular fruit and nut acreage program. An effort began in early February to formulate alternative methods of estimating grape plantings and removals on a more current basis than provided by the regular program. A series of alternative plans were prepared as a basis for discussion with the Wine Advisory Board. They ended up providing half the funds under the matching fund program. These plans ranged from a complete census enumeration of major grape growing counties to a sample mailing and non-response of grape and other fruit operations on the regular program list combined with a non-overlap area sample. Following discussions of these alternatives a general approach was formulated. The survey was to provide reliable estimates of acreage for all grapes, wine grapes, table grapes, raisin grapes, most varieties, and by years of planting on a county level for the major grape growing counties. Also, estimated acreage of varieties by year of planting should be quite precise at the State level.

To develop recommendations needed to determine allocation of resources for the survey, various analyses utilizing orchard record data provided by the regular fruit and nut acreage program were made. Five counties were selected to represent different areas and types of operations. Variances, standard deviations and population coefficients of variation were computed for selected varieties planted in 1968 and in 1969 in the counties. The data were also used to simulate the proportion of annual plantings and total inventory accounted for by operations on the latest orchard record. The simulation was performed by considering 1967 as the year of enumeration. That is, acreage reported by growers having plantings in 1968 and/or 1969 only was assumed to be not covered by the list of growers from a 1967 enumeration. This provided an approximately two year old record, about the average age of the actual orchard record for major grape growing counties. For example in Napa County, there were 466 grape growers. Of these, 24 had no grapes planted prior to 1968, but for Cabernet Savvignon they had 103 out of 289 acres planted in 1968, 49 of 154 acres planted in 1969 and 152 out of a total inventory of 2,006 acres.

Recommendations suggested by the analyses were:

1. A purely sample survey is impractical because planting and removal of grape vines in a given year is a relatively rare occurrence. A mailed request for information to all grape growers is suggested. Variance of 1968 and 1969 plantings for various grape varieties indicates that a very large sample would be required to estimate annual plantings with satisfactory precision.

2. The current orchard record will provide a mailing list of operators and owner-operators which will cover a major portion of new grape plantings and nearly complete coverage of vine removals.
3. A screening survey to identify operations planting grapes which are not on the latest orchard record is necessary. The simulated analysis considering 1967 as the latest year of enumeration indicates that an important portion of new plantings are made by operations not growing grapes in the year of the latest enumeration and thus not on the orchard record for grapes. Most of these operations also are not on the record for any other fruit and nut crops.
4. Because of the failure of some grape operations to respond to a mailed inquiry, a non-response sample survey is required. Every effort should be made to promote a maximum rate of response. The smaller the non-response, the smaller will be the sample required to attain the desired precision in estimating plantings and removals.

The fruit and nut group developed the details aromnd these general recommendations. Lists of growers making recent plantings were provided by a pre-survey of vintners, county commissioners, farm advisors, grape stake suppliers and others. Lists from these sources were reduced to a basic list and checked for duplication. This netted a small, but important, list of growers which was added to the regular orchard record list. Also, all growers identified as having recent plantings were contacted in the non-response follow-up if they did not respond by mail.

All grape growers from the regular fruit and nut acreage list and those gleaned by the screening operation were mailed a questionnaire. An intensive public relations effort in cooperation with the Wine Advisory Board encouraged a large mail response. The non-response survey involved contacting by telephone or field enumeration all non-respondents identified as having recent plantings or having 100 or more grape acres on the record and a one-half sample of the rest.

3. Strawberry Yield Surveys

Strawberry yield survey methodology was reviewed in early 1971. Primary emphasis of the review was upon determination of the relative precision of the grower indicated yield estimate for alternative sampling designs.

The sampling design utilized in past years involved stratification into large and small operations. Large operations were selected with probability one and a systematic sample of small operations arrayed by counties was selected. The systematic sample of small operations provides a type of stratification by counties. The survey is not a probability survey because information is not obtained from non-respondents in either the large or small strata.

The estimates of grower indicated yield are read on various regression charts against Board yield for indications of the State yield. The regression charts are utilized to reduce any consistent bias in grower indicated yield. This bias may be affected by several factors such as; grower yield is for harvested and unharvested portions of the crop, growers tend to underestimate yield, only growers responding are available for use in computing indicated yield and other factors. For the regression charts to remain valid, changes in the methodology for obtaining the estimate of grower indicated yield must be made with caution.

Data from the July 1, 1970 survey were used in an analysis of the precision expected by various types of stratified sampling. For each of eight types of stratification both proportional and optimum sample allocations were developed. For each allocation and stratification combination the standard error and coefficient of variation of estimated strawberry yield was computed for a sample of size 100. The 100 was used only for comparative purposes and ease of calculation. The types of stratification and coefficient of variations were:

None (simple random sample)	- 3.4% (allocation does not apply)
North and South Areas	- both 3.4%
Counties	- both 3.1% (county variances were assumed to be equal)
Varieties	- optimum 3.2% - proportional 3.3%
Varieties and North and South Areas	- both 3.2%
Year of planting	- both 3.4%
Year of planting and North and South Areas	- both 3.4%
Varieties and year of planting	- optimum 3.1% - proportional 3.2%

Recommendation based on the review were:

1. Continue to stratify strawberry growers into large and small groups. Obtain a report for all large operations, if possible.

2. Continue to select a systematic sample of small operations arrayed by counties either within variety strata or for all small operations. Arraying by county provides a type of stratification by counties.
3. Stratify small operations by variety class if the cost in time and other resources is small. Proportional sampling for variety strata will produce a small gain in precision for most years and will provide protection against years when differences between variety yields are great. Optimum allocation is not recommended. It would provide a small gain in precision if the optimum were actually attained. Because estimates from only one survey are available of variances within the variety strata and the response rate may vary between strata, attempted optimum allocation could produce a less precise estimate than proportional allocation.
4. Data for small operations should continue to be obtained only for mail respondents. This is necessary so that regression charts will be valid for the grower indicated estimate. Average yield reported by large operations by mail and phone or other contacts should be compared to determine if there is a significant difference between mail respondents and others in the large stratum.

4. Potato Stocks Surveys

Methods of estimating potato stocks on December, January, February, and March 1st were investigated with emphasis upon problems of non-cooperation in the Tulelake-Butte Valley fall potato area. Of the 129 storage operations in the area in the 1970-71 season, a December stocks report was obtained for only 73. The remaining operations consisted of 42 non-cooperators and 14 who could not be contacted. The sample for the three subsequent surveys was apparently selected from only those who reported in December. An attempt is made to read-out any bias introduced by the method of sample selection for the ratio to capacity estimate. However, the regression charts utilized indicated that the bias has not been consistent from year-to-year.

Potato stock record cards were grouped by type of cooperation and optimum strata boundries were determined for each group based upon storage capacity. Three strata were formed for cooperators and two for non-cooperators. A sample of 50 operations was allocated to the strata in proportion to the product of the number of operations and standard deviation of capacity divided by the square root of cost per report. The principle was to consider obtaining reports from non-cooperators as possible, but more costly than obtaining a report from a cooperator. Cost per report was assumed to be $3 for cooperating storage operations, $60 for smaller non-cooperators and $50 for the larger non-cooperators. This procedure indicated a sample of eight non-cooperators and a coefficient of variation of 4.1 percent for the direct expansion estimate of capacity. Precision of estimated stocks in each of the four surveys should be similar to that of estimated capacity inasmuch as (1)stocks are strongly related to capacity and (2) stratification by capacity, groups data on stocks in relatively homogeneous strata.

This exercise using the previous year's data provided an illustration of three alternative methods. All contained the concept of obtaining stock data for operations classed as non-cooperators, but recognized the additional cost of obtaining that data. The plan selected by the vegetable group was to completely enumerate all cooperating storage operations for the December 1 survey and to sample both cooperating and non-cooperating storages with known probabilities in January, February, and March. Information on ownership, tenure, management, capacity, and future likelihood of cooperation had to be ascertained for each storage operation before a probability sample could be selected.

The vegetable group visited the area in late October to obtain information for most non-cooperators and obtained information on all remaining operations while conducting the December 1st survey. The information was utilized to stratify the population of storage operations, allocate the sample and compute the expected precision of estimating total capacity. This was done in the same manner as the illustration on 1970-71 data. The coefficient of variation of the direct expansion estimate of capacity was 3.7 percent.

The vegetable group did an excellent job obtaining information on the 1971-72 storages and also reduced the number of non-cooperating operations. This resulted in fewer non-cooperating storages being sampled than was anticipated. The attitude toward the concept of obtaining good reports for the sampled non-cooperators was reasonably optimistic.

I was able to assist with field work for the January 1st survey. This was the first survey sampling non-cooperators. It was a very worthwhile experience and provided an opportunity to further illustrate the feasibility of the methodology employed. After some explanation of the job we were trying to do, listening with real interest to their complaints and impressing upon them our desire that they assist us in making the best estimate we could; half of the non-cooperators sampled gave a willing report. The other half received the attention corresponding to the cost assumptions. We believe a good report of their potatoes in storage was also obtained.

The direct expansion and ratio expansion for January 1 were quite consistent. The estimated coefficient of variation of the direct expansion estimate was 5.7 percent. The survey data supported five estimates of 50,000 cwt. increments within a one standard error interval and outshipments check data pointed to one of those. We are of course interested in how the February and March surveys will perform. Aside from the survey estimates, it will be of interest how contacting non-cooperators affects their longer term performance.

5. Honey Production Estimates

Because of changes in the honey estimating program, stratification, sample allocation and expected precision of estimated number of colonies were investigated. The cumulative square root of frequency procedure indicated optimum boundaries for five strata based upon recorded number of colonies. Stratification was subject to the constraint that beekeepers with 300 or more colonies (commercial) be in different strata from those with fewer colonies. Two strata were defined for the smaller operations and three for commercial beekeepers.

Expected variance, standard error and coefficient of variation were computed for the direct expansion estimate of total colonies for each type of allocation and sample size. The estimates of precision were developed for all colonies, commercial colonies and others. For proportional allocation to individual strata samples of size 500 and 400 were considered. For optimum allocation samples of 500, 400, and 300 were used. The estimated coefficients of variation ranged from 1.3 to 6.6 percent for estimated total colonies, 0.7 percent to 7.5 percent for commercial beekeeper's colonies and 7.5 to 13.5 percent for noncommercial group. Optimum allocation was superior to proportional, but sample size was largely dependent on the time available to collect the data.

6. Field Crop Acreage Research Project

The field crop acreage research project was initiated to study the utility of a list as a sampling frame for estimating field crop acreages. The objective of the project is to study step-by-step the utility of the California Agricultural Disability and Unemployment Insurance (DI-UI) list as a sampling frame for major field crop surveys. The project consists of three phases:

Phase I. Identify the agricultural coverage of the DI-UI list as represented in the 1971 June Enumerative Survey.

Phase II. Analyze the relationship of various supplemental information on the DI-UI list to field crop data for crops with reasonably complete coverage.

Phase III. Study and evaluate the efficiency of various sample designs utilizing information from the list frame. Establish procedures for and conduct a pilot survey to test the efficiency and feasibility of the design in a major field crop survey (March Intentions, June Acreage, Early A&P and Late A&P).

To investigate the possibility of conducting probability field crop acreage surveys, it is first necessary to determine the availability of a suitable sampling frame. A suitable frame must be reasonably complete in its coverage of the population of interest. The level of completeness of a frame may indicate any of three conclusions: (1) the frame is not sufficiently complete, (2) the frame is sufficiently complete to provide an efficient sampling design when coupled with an area frame in a multiple frame type survey or (3) the single frame provides nearly complete coverage of the population of interest. To minimize cost, it is desirable that the frame be one for which the majority of data can be obtained by mail.

The DI-UI list seems to be a good candidate frame for estimating crop acreages. The insurance code requires all agricultural employers to file for disability insurance for their employees if they paid wages of more than $100 in any calendar quarter. They may, also, elect unemployment insurance coverage. Once an employer has filed, he must report quarterly for the current calendar year and the following calendar year even though wages may not exceed $100 again. The quarterly report includes all wages paid and workers employed during pay periods ending nearest the 12th of each month. In addition to employment data, the list has the advantage of being classified by industry codes (field crops; fruit tree and vegetables; livestock, general farms and so-forth).

To provide a preliminary evaluation of the list, a special data listing of all resident farm operators with cotton in the 1971 June Enumerative Survey was obtained. Operations on the list were identified in four classes based on June Enumerative questionnaire answers about DI-UI list membership. Farm expansions of cotton planted and operations planting cotton were computed for each class. Operations filing a DI-UI application form accounted for 88 percent of acreage and 68 percent of cotton operations. Operations not filing, but paying more than $100 during a calendar quarter (legally they should have filed) represented 8 percent of acreage and 7 percent of operations, the remainder was accounted for by operations hiring agricultural workers, but paying less than $100 and not filing and by operations not hiring workers. Based upon uncertain list membership for those reporting they paid more than $100 but did not file, completeness of the list is 88 to 96 percent for cotton acreage and 68 to 75 percent for operations planting cotton.

The first phase of the project is underway. A program has been written to identify 1971 June Enumerative Input I records by classes based upon answers recorded on Input II punch cards. The classes identify operations as those (1) filing a DI-UI application, (2) not filing, but paying more than $100 during a calendar quarter, (3) not filing or paying more than $100, but hiring agricultural workers, and (4) those not hiring agricultural workers. Another program is now being written to expand all Input I commodity items by each class. Closed segment items will be adjusted to the weighted farm level before expansion. The completion of Phase I will yield an estimated range of completeness for each Input I commodity item as of June 1971. This will provide a basis for evaluating the utility of the DI-UI list as a sampling frame with respect to various crops and other commodities and survey groupings.

7. Grape Objective Measurement Forecast Survey

A review of the grape objective measurement forecast survey was begun with a review of the basic literative on grape growth and cultural practices. This was followed by some preliminary study of the current status of the grape yield forecasting program. The performance of the program as evaluated against final Board yield for the years 1940-58 and 1959-70, indicated some improvement from the objective measurement indications which began to affect statist estimates in 1959. The improvement, however, was not so marked as to make one complacent about the need to improve the accuracy of the estimates. Correlation coefficients of grower reported percent of full crop and Board yield and production for the 1959-69 period indicated that the grower survey is of considerable value.

Three areas of potential improvement of the forecasts were chosen as having the greatest promise. Briefly described they are:

1. Heat summation data may be useful in adjusting objective measurement data for year-to-year variation in timing of the survey relative to the stage of development of the crop.

2. Grower-reported yield per acre could be very useful in the evaluation of new objective measurements for which a historical series has not been built up. This would facilitate the testing of potentially valuable measurements without burdening the survey with relatively useless observations for a series of years.

3. Because southern California (mainly Coachella area) table and raisin grapes are excluded from the objective measurement survey sample, but the area's production and acreage are included in Board yield, the regressions may be less precise than they could be and might be biased. Coachella area table and raisin grape acreage and production should be excluded from the computation of yield for use in the regressions.

Progress on this review has been limited to assembling basic data for 14 items from which a total of approximately 30 items potentially related to yield will be developed. One of these items is heat summation. Heat summation is defined as the sum of mean temperature minus 50°F. for each day, that mean temperature exceeds 50°F. Approximately 20 weather stations have been selected in grape growing areas to reflect (hopefully) crop development on selected dates. Heat summation data will be associated with each block based on area and the data block data was obtained. This will provide an average of heat summation "degree growing days" for all sample blocks at the time of the survey.

After computing heat summation data for 1971 for most stations, it appears that for grapes heat summation data should be started March 1. This avoids picking up earlier warm periods which are not sustained enough to start development. Ideally, one might like to begin cumulating heat summation at the date of blooming. If this heat summation approach proves useful, we will be interested in its application to other objective measurement surveys.

8. Earth Resources Technology Satellite (ERTS) and the California Proposal

In early 1971, a request for data and imagery from the simulated ERTS high flight and ERTS satellite program was made through the California Department of Agriculture. The project that we initiated at that time was to be a study of the utility of ERTS photography for agricultural uses in cooperation with two other Bureaus and County Agricultural Commissioners. Potential uses of ERTS data and imagery to be investigated were:

1. General uses.
2. Identification of land use.
3. Measurement of identified areas.
4. Forecast of crop yield potential.
5. Evaluation of plant damage due to episodes and epidemics.

All of these uses involve some comparison of ERTS data and ground data.

Subsequently, our proposal was included in a State of California proposal submitted to NASA in late April. Potential uses to be investigated were not changed, but imagery requested was generalized to the needs of the nine departments involved in the proposal. To date we have not been notified whether the proposal has been approved by NASA.