Point Sampling Surveys for Potato Acreage in Colorado's San Luis Valley

By H. F. Huddleston

DURING THE PAST FEW years considerable attention has been given to the collection of agricultural data by probability sampling.

Enumerative surveys employing a large group of trained interviewers often are used to obtain such data. Multiple frame surveys using lists of specialized producers to supplement an area frame may be employed, or several lists may be used to screen the population for the elements with the desired characteristics. In some situations, less sophisticated probability techniques requiring a relatively small group of trained people may offer possibilities of rapid and efficient execution in the field.

If there is a problem of available resources for survey work, the technique of point sampling may provide an alternative which should be considered. Point sampling is a method of observing the use of very small units of areas of equal size whose locations are determined by random points. The sizes of the fields in which the random points fall are immaterial.

This paper gives an example of the application of the technique with the related costs and variances. It presents a brief summary of the survey methodology used during the first 2 years of a project in Colorado. The project was initiated in June 1964 in response to industry desires for an early season forecast of potato acreage for the San Luis Valley in Colorado. The statistician in charge of the Colorado office of the Statistical Reporting Service, R. S. Overton, and his staff directed the survey in the field and assembled source materials for the frame. The county offices of the Agricultural Stabilization and Conservation Service were the source for aerial photographic coverage of the sample units selected for the survey.

The Sampling Frame

There was a need for a sampling procedure which would give an objective and independent estimate of potato acreage in late June. Three sampling frames appeared possible:

1. A list of growers who sold or stored potatoes in the preceding year could be assembled.
2. The materials in the master sample of agriculture could be used.
3. A special area frame could be constructed.

The third frame was considered the most suitable. The list frame in No. 1 was not considered since it might be biased by being out of date, or otherwise incomplete. The frame in No. 2 was not considered efficient because the total land area seemed too large and would result in a large sampling error.

The use of a two-stage point sampling scheme offered the following advantages:

1. A relatively cheap method of direct observation of land use compared with interviewing producers; fastest means of data collection.
2. Objectivity and freedom from certain grower response biases commonly associated with the enumerative or self-enumeration surveys.
3. A ready sample of potato farms and fields for yield estimates and variety classification later in season.
4. Could be executed by a small number of people already trained and employed.

The sampling frame used was an area frame constructed in 1964 to contain the potato acreage in five counties comprising the San Luis Valley. The land area in the frame, totaling 372,800 acres, was subdivided into 563 sampling units (S.U.) of nearly equal land area. The 563 sampling units were grouped into three strata based on contiguous land area and estimated percentage of total land planted to potatoes. The 1964 frame composition by strata was as follows:

Stratum	Land area (acres)	Number of frame units
1.	191,616	289
2.	132,800	201
3.	48,384	73
Total.	372,800	563

After the 1964 surveys were completed, an analysis and evaluation of the data suggested several modifications in the strata as originally constructed: (1) Several large "islands" of noncultivated land could be excluded from the frame since potatoes were not produced on this type of land, and (2) isolines based on the percentage of points in potatoes indicated different stratum boundaries would reduce the within-stratum variability. The modified frame composition was made up as follows:

Stratum	Land area (acres)	Number of frame units
1.	69,120	108
2.	121,984	184
3.	90,432	139
4.	32,960	49
Total	314,496	480

Sample Design and Estimates

Two estimators were derived, but the results of the 1964 and 1965 surveys (see tables, p. 3) correspond to the first estimator here:

$$\text{(1)} \quad (\text{Potato acres})_h = P_h \cdot (\text{Land area})_h$$

$$\text{(2)} \quad (\text{Potato acres})_h = \frac{N_h}{n_h} \; (53.33) \sum_{i=1}^{n_h} Y_{hi}$$

The two estimators gave nearly identical estimates of potato acreage for the total population. This is to be expected if the maps used to determine measured land area in the frame and the overlay grid used to locate points on the aerial photographs all had correct scales and the sampling procedures were properly carried out.

The within-stratum variances for the percentage of points in potatoes were computed using the following formula:

$$S^2_{ph.} = \left(\frac{N_h - n_h}{N_h \, n_h \, (n_h - 1) \, \overline{X}^2_{h.}} \right) \times$$

$$\left(\sum_{i=1}^{n_h} Y^2_{hi} + P^2_{h.} \sum_{i=1}^{n_h} X^2_{hi} - 2 P_{h.} \sum_{i=1}^{n_h} X_{hi} Y_{hi} \right)$$

where

$\overline{X}_{h.}$ = average number of points per S.U. in hth stratum

X_{hi} = number of points for ith S.U. in hth stratum

Y_{hi} = number of points found in potatoes for ith S.U. in hth stratum

$P_{h.}$ = percentage of points in potatoes $\overline{Y}_{h.} \div \overline{X}_{h.}$

N_h = number of S.U. in hth stratum

n_h = number of S.U. sampled in hth stratum

The standard error for the acreage estimate is $S_{Ah} = S_{ph} \cdot (\text{Land area})_{h.}$

To insure the completeness of the acreage estimates for the entire valley, growers who had land outside the frame area and who had previously grown potatoes were enumerated each year. The enumeration in 1965 uncovered an additional 400 acres. This acreage was on marginal potato land and was not likely to increase substantially unless new land was brought into cultivation. But to prepare for such increases, provision was made for new land to enter into the frame each year and to be sampled at the same rate as the previous sampling units in the appropriate strata.

The use of a list frame for stratum 4 appeared desirable if estimates were to be published by strata, since relatively few growers were involved and their acreage could be completely enumerated each year. However, this was not expected to reduce the sampling error for the entire valley. The sampling plan adopted was to use the stratified two-stage point sample scheme for strata 1, 2, and 3, with a list frame for stratum 4.

A stratified sample of 100 first-stage sampling units (area segments) was used in 1964 with points as secondary units located within the segments at a rate of one point per 53 acres.

Results of the 1964 survey are shown here:

Stratum	Number of units sampled	Estimated potato acreage	Standard error of acres
1.	52	22,592	3,301
2.	35	8,433	2,336
3.	13	1,210	531
Total. . . .	100	32,235	4,079 (12.7%)

Results of the 1965 survey using a modified frame, optimum number of points, and an independent selection of first-stage units are shown below:

Stratum	Number of units sampled	Estimated potato acreage	Standard error of acres
1.	27	17,156	1,544
2.	46	9,088	1,181
3.	34	8,573	1,526
4.	13	517	201
Total . . .	120	35,334	2,480 (7.0%)

Two-Stage Point Selection and Cost Per Sampling Unit

A systematic selection of first-stage sampling units was made in each stratum using a random start. The sampling units in the frame were listed in a serpentine fashion from east to west and west to east, starting in the northeastern corner of the stratum. To locate the points within selected segments, a transparent plastic overlay with a systematic pattern of small holes was used. A random starting corner was used for each sampling unit. This consisted of selecting two random coordinates between zero and D where D represents the distance between points on the overlay, and matching this point with the point furthest north and west in the segment on the photograph. A narrow pen with a fine point was inserted in the small holes in the overlay to ink the points permanently on the photographs for field inspection. The points were located at a rate of one point per 53 acres in 1964 with the total number of points being a variable which ranged over rather narrow limits because of the nearly equal segment size. This procedure resulted in an average of 12.6 points per sampling unit in 1964.

An analysis of the 1964 data on variability and costs indicated the optimum number of points per S.U. should be approximately 14 in strata 1, 2, and 3, and 30 or more in stratum 4. These modifications were made for the 1965 survey. The average cost components per sampling unit in 1964 were: (1) Between S.U., \$3.11; (2) within S.U., \$1.55; and (3) field supervision, \$2.19. Costs were based on payment of 9¢ a mile for transportation and salary of \$3 an hour. The technique for determining the optimum cluster size can be found in Hendricks (see References) or other available textbooks. The relationship used to determine the optimum number of points was:

$$m_{opt} = \left(\frac{\sigma_W^2}{\sigma_B^2} \cdot \frac{C_B}{C_W / \overline{m}} \right)^{\frac{1}{2}}$$

where

C_B = cost between S.U.
C_W = cost within S.U.
$\overline{m}$ = average number of points per S.U.

and σ_W and σ_B are the variance components found within and between sampling units in the same stratum, using analysis of variance of the variable P_{hi} (i.e., $Y_{hi} \div X_{hi}$). The pooled sample estimates of σ_W^2 and σ_B^2 from the 1964 data were 0.07013 and 0.00922.

A breakdown of the cost per sampling unit follows:

Between sampling units--C_B	
Mileage--19 miles at 9¢	$1.71
Salary--28 minutes at 5¢	1.40
	$3.11
Within sampling units--C_w	
Mileage--5 miles at 9¢.............	$ .45
Salary--22 minutes at 5¢	1.10
	$1.55
Field supervision	
Mileage--15 miles at 9¢	$1.30
Salary--16 minutes at 5¢	.80
Per diem--0.01 day at $9..........	.09
	$2.19
Total field costs...................	$6.85

Field work for the survey required use of three people for three days.

In most cases the points could be inspected by driving a car along the field edge. However, inspection of the land use beneath some of the points required the technician to walk as much as half a mile.

Reinspection of the sample points later in the season for variety and yield work revealed only one point incorrectly classified.

References

1. Kelly, B. W., "Probability Sampling in Collecting Farm Data", Jour. Farm Econ., Vol. 45, No. 5, Dec. 1963, p. 1515-1520.
2. Hartley, H. O., "Multiple Frame Surveys," Social Section, Amer. Statis. Assoc., Proc., 1962, p. 201-203.
3. Hendricks, W. A., The Mathematical Theory of Sampling, Scarecrow Press, New Brunswick, N.J., 1956, p. 198-199, 250-255.