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

by H. F. Huddleston

Introduction

The work was initiated in June of 1964 in response to industry desires for an early season forecast of potato acreage for the valley. This paper presents a brief summary of the survey methodology used during the first two years. The statistician in charge of the Colorado office, R. S. Overton, and his staff directed the project in the field and assembled source materials for the frame. The county ASCS offices provided aerial photographs of the sample units used in the survey.

The need for a sampling procedure which would give objective and independent estimate of potato acreage in late June was desired. Several sampling schemes appeared possible:

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

The later frame was considered most suitable since the frame in (2) was not considered efficient since the total land area was considered too large and would result in a large sampling error. The list frame in (1) might be limited because of being outdated and incomplete.

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

1. Observation of land use was relatively cheap as compared to interviews, surveys, and the fastest means of data collection.
2. Objective and free of certain grower response biases thought to be associated with the enumerative (or self-enumeration) type of survey.
3. Provided a ready sample of potato fields for yield work and variety determinations later in season.

Sampling Frame:

The sampling frame used was an area frame constructed in 1964 to contain the cultivated land in 5 counties comprising the San Luis Valley. The land area in the frame totaled 372,800 acres which was subdivided into 563 sampling units of nearly equal land area. The 563 sampling units were grouped into 3 strata based on contiguous land area and the estimated relative fraction of total land planted to potatoes. The 1964 frame composition by strata was as follows:

Strata	Land Area (acres)	Number Sample Units
1	191,616	289
2	132,800	201
3	48,384	73
Total	372,800	563

Following the 1964 surveys an analysis and evaluation of the data suggested several modification in the strata as originally constructed: (1) Several large areas of land (i.e., islands of non-cultured land) could be excluded from the frame since potatoes were not produced on this type of land, and (2) Isolines based on fraction of points in potatoes indicated different strata boundaries would reduce the within strata variability. The modified frame composition was made up as follows:

Strata	Land Area (acres)	Number Sample Units
1	69,120	108
2	121,984	184
3	90,432	139
4	32,960	49
Total	314,496	480

Two Stage Point Selection Procedure:

A systematic selection of sampling units using a random start in each strata (the sampling units in the frame were listed in a serpentine fashion from east to west starting in the northern most corner of the strata). Within each selected sampling unit a predetermined systematic pattern of points using a random corner location for each sampling unit was used to spot points at a rate of one point per 53 acres. This resulted in an average of 12.6 points per sampling unit.

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, 3, and 30 or more in strata 4. These modifications were made for the 1965 survey. The average cost components per sampling unit in 1964 were: (1) Between S. U. costs \$3.11, (2) Costs within S. U. \$1.55, and (3) Field supervision costs \$2.19. Costs were based on payment of .09¢ a mile for transportation and salary of \$3.00 an hour. The technique for determining the optimum number of points can be found in Hendricks[1] or other available texts.

$$m_{opt} = \left[\frac{\sigma_w^2}{\sigma_B^2} \cdot \frac{C_B}{C_w/\bar{m}} \right]^{\frac{1}{2}}$$

where

C_B = cost between S. U.

C_w = cost within S. U.

$\bar{m}$ = 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 strata 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 .07013 and .00922.

Table 1 Cost per Sampling Unit

Between Sampling Units - C_1	
Mileage - 19 miles at 9¢	$1.71
Salary - 28 minutes at 5¢	1.40
	$3.11
Cost within Sampling Units - C_2	
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 - .01 day at $9	.09
	$2.19
Total Field Costs	$6.85

Sample Design and Estimates:

A stratified sam ple of 100 first stage sampling units was used in 1964. The sample allocation, the estimated potato acreage and standard errors by strata for 1964 are shown in the Table 2 below:

Table 2

Strata	No. 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%)

The corresponding information for 1965 based on the modified frame, optimum number points and an independent selection of first stage units are shown in Table 3 below:

Table 3

Strata	No. Sampling Units	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%)

The within strata variances for the fraction of points in potatoes were computed using the following formula:

$$S^2_{Ph.} = \frac{N_h - n_h}{N_h\, n_h\, (n_h - 1)\, \bar{X}_h^2} \left(\sum_{i=1}^{n_h} y^2_{hi} + P^2_{h.} \sum_{i=1}^{n_h} X^2_{hi} - 2\, P_{h.} \sum_{i=1}^{n_h} X_{hi}\, y_{hi} \right)$$

where

$\bar{X}_h$ = average number of points per S. U. in h^{th} strata

X_{hi} = number of points for i th S. U. in h^{th} strata

Y_{hi} = number of points found in potatoes for i th S. U. in h^{th} strata

P_h = fraction of points in potatoes $\bar{Y}_h \div \bar{X}_h$

N_h = number of S. U. in h th strata

n_h = number of S. U. sampled in h th strata

The standard error for the acreage estimate being $S_{A_h} = S_{p_h} \cdot (\text{Land area})_h$.

Two estimators were derived, but the results in Tables 2 and 3 correspond to the first estimator below:

(1) $(\text{Potato Acres})_h = P_h \cdot (\text{Land Area})_h$

(2) $(\text{Potato Acres})_h = \frac{N_h}{m_h} (53.33) \sum_{i=1}^{n_h} y_{hi}$

The two estimators gave nearly identical estimates of potato acre for the total population. This is to be expected if the maps used to determine measured land area in the frame and the point grid used to locate points on photos had correct scale and the sampling procedures was properly carried out.

To insure the completeness of the acreage estimates for the entire valley the acreage for growers who had land outside the frame area and who had grown potatoes was enumerated each year. This additional acreage amounted to 400 acres in 1965. This acreage was on marginal potato land and was not likely to increase substantially unless new land was brought into cultivation. To guard against this possibility, provision is made for new land to entered into the frame each year and be sampled at the same rate as the previous sampling units in the appropriate strata.

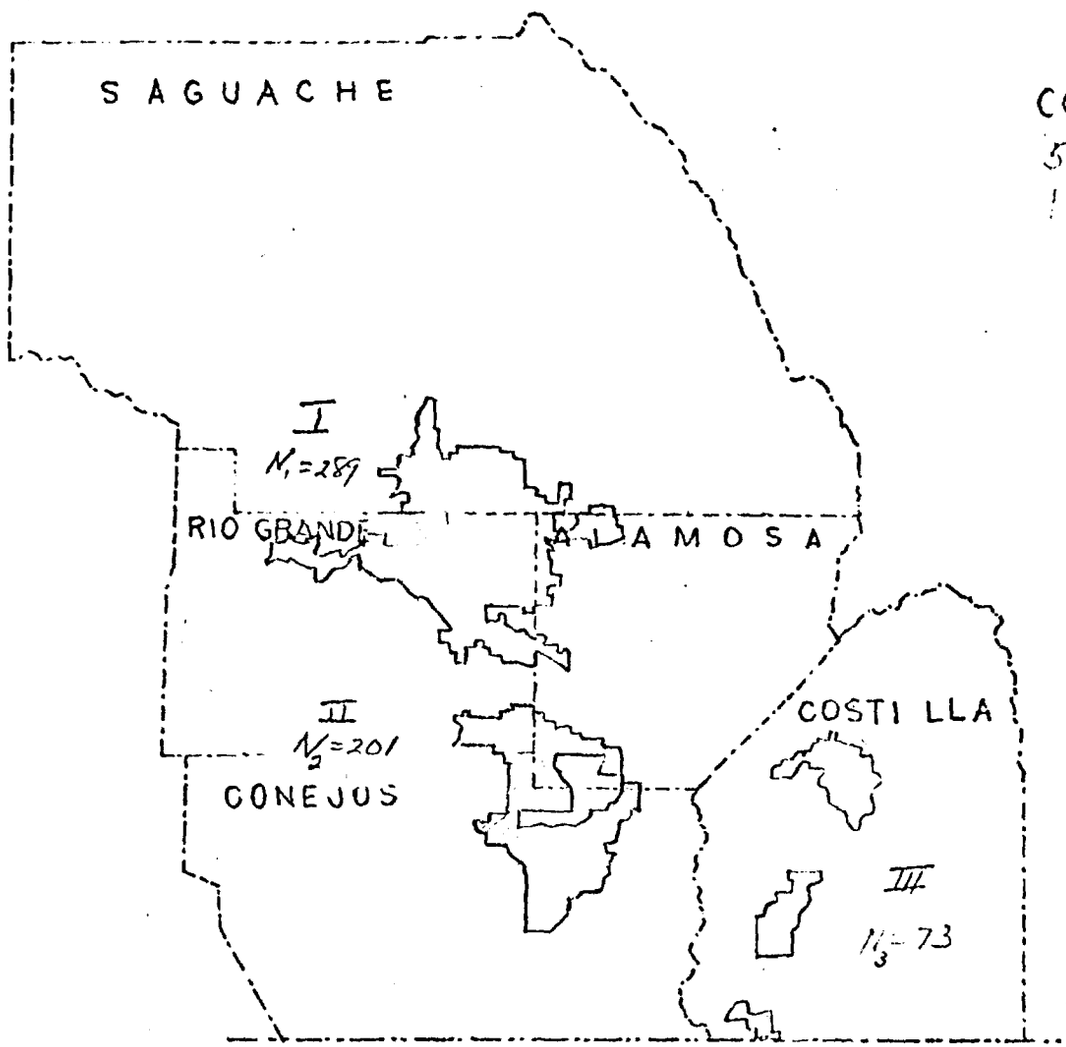
The use of a list frame for Strata 4 appeared desirable if estimates were to be published by strata since relatively few growers were involved and their acreage could be enumerated each year. The sampling plan adopted was to use the stratified two stage point scheme for Stratas 1, 2, and 3 with a list frame for Strata 4.

References:

Hendricks, W. A., The Mathematical Theory of Sampling, Chapter 8, Scarecrow Press 1956

Appendix: (1) Map of 5 county area and 1964 strata

(2) Analysis of variance procedure for estimation of variance components.



COLORADO
5-COUNTY
[illegible]
AREA
1961

ANALYSIS OF VARIANCE TABLE

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Mean Square provides Estimate of:
Between Strata	$BSS = \sum_{i=1}^{h} K_h (\bar{P}_{i\cdot} - \bar{P}_{\cdot\cdot})^2$	$h-1$	$\frac{BSS}{h-1}$	$\sigma_w^2 + \bar{X}_0' \sigma_B^2 + K_0 \sigma_S^2$
Between S.U. Within Strata	$BWSS = \sum_{i=1}^{h} \left[\sum_{j=1}^{m_h} X_{ij} (P_{ij} - \bar{P}_{i\cdot})^2 \right]$	$\sum_{i=1}^{h} (m_h - 1)$	$\frac{BWSS}{\sum_{i=1}^{h} (m_h - 1)}$	$\sigma_w^2 + \bar{X}_0 \sigma_B^2$
Within S.U.	$WSS = \sum_{i=1}^{h} \left[\sum_{j=1}^{m_h} X_{ij} P_{ij} (1 - P_{ij}) \right]$	$\sum_{i=1}^{h} (K_h - m_h)$	$\frac{WSS}{\sum_{i=1}^{h} (K_h - m_h)}$	σ_w^2

1. Definition of Variables:

X_{hi} and P_{hi} are defined for individual sampling units (see earlier discussion)

$\bar{X}_0$ = average number of points per S.U. within strata

$\bar{X}_0'$ = " " " " " between strata

K_0 = " " " " " per strata

$$\bar{X}_0 = \frac{1}{\sum_{i=1}^{h} (m_h - 1)} \left[\sum_{i=1}^{h} \left(\sum_{j=1}^{m_h} X_{ij} \right) - \frac{\sum_{j=1}^{m_1} X_{1j}^2}{\sum_{j=1}^{m_1} X_{1j}} - \frac{\sum_{j=1}^{m_2} X_{2j}^2}{\sum_{j=1}^{m_2} X_{2j}} - \frac{\sum_{j=1}^{m_3} X_{3j}^2}{\sum_{j=1}^{m_3} X_{3j}} \right]$$

$$\bar{X}_0' = \frac{1}{h-1} \left[\frac{\sum_{j=1}^{m_1} X_{1j}^2}{\sum_{j=1}^{m_1} X_{1j}} + \frac{\sum_{j=1}^{m_2} X_{2j}^2}{\sum_{j=1}^{m_2} X_{2j}} + \frac{\sum_{j=1}^{m_3} X_{3j}^2}{\sum_{j=1}^{m_3} X_{3j}} - \frac{\sum_{i=1}^{h} \left(\sum_{j=1}^{m_h} X_{ij}^2 \right)}{\sum_{i=1}^{h} \left(\sum_{j=1}^{m_h} X_{ij} \right)} \right]$$

$$K_0 = \frac{1}{h-1} \left[K_1 + K_2 + K_3 - \frac{K_1^2 + K_2^2 + K_3^2}{K_1 + K_2 + K_3} \right]$$

2. Find Variance Components:

Solve for σ_B^2 and σ_S^2 after σ_w^2, $\bar{X}_0$, $\bar{X}_0'$, and K_0 have been found.