ESTIMATING DRY BEAN ACREAGE IN MICHIGAN: SECOND YEAR RESULTS, by Ron Fecso and Jeff Geuder, Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture. SRS Staff Report No. AGES831020.

ABSTRACT

The dry bean survey again provided acreage estimates which were useful to the State Statistical Office and the industry cooperators. A second year of data gave the opportunity to critique the survey and to further evaluate survey costs and auxiliary data. Changes to be made in the 1983 survey as well as proposals for future research are presented. An improved late season estimator for other major crops is also developed.

Keywords: Area frame, dry bean estimates, survey design

ACKNOWLEDGMENTS

The quality of the survey work and the enthusiasm of the personnel in the Michigan State Statistical Office is appreciated. The authors wish to thank Carol House for a critical review of the report. The word processing efforts of Karen Watkins are also appreciated.

Washington, D.C. October 1983

CONTENTS

SUMMARY

Dry bean surveys were conducted in Michigan in July 1981 and 1982 to obtain more precise acreage estimates than were available from the June Enumerative Survey (JES). The area frame used in the JES does not efficiently estimate a specialized crop such as dry beans. Also, the JES is conducted in late May and early June, before much of the dry bean crop is planted in Michigan. To date, the dry bean estimate made from the JES (which includes a significant amount of intentions rather than actual plantings) has exhibited a downward bias.

The area frame used in the dry bean survey (DBS) was constructed specifically for estimating acreage in a 16 county area around Lake Huron, an area which produced over 90 percent of the dry beans in Michigan. The direct expansion estimate (closed segment approach) from the DBS had a coefficient of variation (C.V.) of about 8 percent. This compares with JES C.V.'s of 12.99 percent in 1981 and 14.8 percent in 1982.

A weighted estimate was obtained from the 1981 DBS. This estimate was significantly different from the closed estimate, a fact which further substantiates previous work showing that a weighted estimate using total land for the weight is biased. The weighted estimate was not used in 1982, a fact which contributed to a 35 percent reduction in enumeration expenses before allowing for inflation.

An attempt was made, in both years, to use available data (such as soils maps and county estimates) to assign auxiliary data to frame units. The auxiliary data was then used to compute regression estimates of dry bean acreage. Regression coefficients were preassigned to produce an unbiased estimate. These regression estimates in both years were only slightly more precise than the direct expansion estimates. This was somewhat disappointing; however, there are potential gains to be made in this area and futher research is recommended.

This paper presents an examination of alternative survey designs and costs. This is intended to answer some questions concerning the report on the first dry bean survey and to serve as an aid in future specialty survey designs. The paper also proposes a late season estimator for other crops which should be more precise than the JES direct expansion. Suggestions are made for future research concerning survey costs, use of soils, stratification, and estimation.

ESTIMATING DRY BEAN ACREAGE IN MICHIGAN: SECOND YEAR RESULTS

Ron Fecso
Jeff Geuder

INTRODUCTION

In 1981, the dry bean industry had a formal export agreement with Mexico. This agreement was not renewed in 1982, and the reduced exports were expected to result in a significantly lower acreage of dry beans in the 16 county area encompassed by the dry bean survey. A complete description of the special sampling frame used in the Dry Bean Survey (DBS) is contained in an earlier report. (1) 1/ This 16 county area produces over 30 percent of the nation's dry beans and 80 percent of the navy beans. The economic impact of the production in this region makes it important to produce a reliable estimate of dry bean acreage for the area.

The survey design used in 1982 was unchanged from the initial design. There was no rotation or reallocation of segments, and there was no restratification (which had been considered after the first year). Since the design was unchanged and a large acreage change was expected, the effectiveness of the special frame for dry beans could be evaluated.

The analysis of the 1981 DBS indicated that the weighted estimate was biased, a result consistent with other research. Because of this bias and the higher enumeration costs associated with the weighted estimator, the information needed for this estimator was not collected in 1982. The closed and regression estimators were used again, as was a ratio estimator applied to the previous year's Crop Reporting Board estimate of planted acres. A regression estimator using a new auxiliary variable was tried with limited success.

RESULTS

The Michigan SSO edited the survey data using the SRS Generalized Edit System. The edited data tape was then accessed by the Sampling Frame Development Section to be summarized using the Area Frame Analysis Package (4). Survey results are discussed below for each of the estimates computed.

1/ Underscored numbers in parentheses refer to literature cited in the references.

Direct Expansion Estimates

Closed estimates were computed for dry beans varieties in the 16 county survey area. Table 1 summarizes the direct expansion estimates and associated standard errors by survey year and variety. Black Turtle beans showed the largest change - a decrease of over 100,000 acres. The total acreage for all varieties decreased almost 90,000 acres.

Other Estimates

In 1981, the auxiliary variable (expected acres of dry beans) was based on an average soil type in each count unit. In some strata, this approach did not work well. There were some count units which contained primarily soil unsuited for dry beans and a small amount of "good" soil. The segment selected from some of these count units was in the area with the good soil, and contained a large amount of dry beans. In these strata, a new approach was tried in 1982. The auxiliary variable was based on the "best" soil type in the entire count unit. "Best" refers to the soil type best suited to dry beans. The regression estimate and difference estimate were computed in 1982 (Table 2) using this new auxiliary variable. (The difference estimate is equivalent to a regression estimate using a regression coefficient of 1.0 in all strata. The regression coefficients which were preassigned for each stratum are given in an Appendix to the first report. They were unchanged in 1982). These apply only to the 16-county area, not the entire state.

Table 1 - Direct expansion estimates for dry beans by variety and year of survey.

Variety	1981 Estimate	1981 Standard error	1982 Estimate	1982 Standard error
	(1,000 acres)			
Navy	427.1	43.6	456.5	36.8
Dark Kidney	8.7	3.9	17.3	5.6
Light Kidney	8.9	4.7	4.9	1.9
Cranberry	19.4	6.6	21.3	8.1
Yellow Eye	1.7	1.2	.6	0.6
Pinto	16.4	4.7	10.7	5.5
Black Turtle	122.2	18.6	4.6	2.3
Other	7.2	1.0	6.6	3.9
Total	611.6	50.2	522.4	41.2

Table 2 - Direct expansion, regression, difference, and ratio estimates, 1982.

Estimator	Estimate 1/	Standard error	Coefficient of variation
	(1,000 acres)		(percent)
Direct Expansion	522.4	41.2	7.9
Regression	509.9	39.4	7.7
Difference	497.8	39.9	8.0
Ratio 2/	540.4	38.0	7.0

1/ Estimates apply to 16-county area.
2/ Ratio applied to 16-county proportion of 1981 Board Estimate.

The regression estimate using coefficients computed from the sample data was not used as a survey indication, due to the possibility of large bias, as pointed out in the earlier report (1). However, the estimate was computed as a research tool to evaluate the potential of this approach. The estimate was slightly higher than the unbiased regression estimate, and had a lower standard error. With the small sample sizes in the various strata, it was very difficult to predict the optimum values of the coefficients. However, the fact that the estimate had a higher level of precision shows that the regression estimator has potential. Its precision may be improved by futher research on optimizing the preassigned regression coefficients.

The ratio estimate could be computed because the entire sample was used in both survey years. This ratio estimate was the most useful to the Michigan SSO in estimating planted acreage for 1982. As was pointed out in the report on the 1981 survey, the JES is not very well suited to estimating dry bean acreage, primarily because it is conducted before much of the crop is planted, and changes in intentions are not reflected in the estimate. The SSO was also expecting a large decrease in acreage due to the discontinuance of the export agreement. The ratio estimate was computed as the product of the ratio of planted acreage in 1982 to that in 1981 and the 16-county proportion of the 1981 Crop Reporting Board estimate of 650,000 acres planted. The 16-county proportion was determined by using the ratio of the direct expansion estimate in the 16 counties for 1981 to the direct expansion for the entire state, as reported in (1).

EVALUATION OF THE SURVEY DESIGN

The dry bean survey has been well received by the state and industry cooperators. The direct expansion estimates from the survey have been only slightly more precise than the estimate that could be obtained with a comparable sample size in the JES. However, the DBS does provide varietal estimates (particularly the Navy and "all other" breakdown) which the JES does not. This is very important to the Michigan SSO. This section addresses the strengths and weaknesses of the techniques used in this research-oriented survey design thus providing the foundation for future research.

Construction of Frame Units

As discussed in the first report, frame units were constructed such that within the available boundaries, the land was homogeneous with respect to the percentage of cultivation, woods, water, waste, urban development, and to the extent possible, soil type. These variables and a geographically defined "possibility of dry beans" were added to each primary sampling unit in the frame and used to classify the frame units into strata. This method differs from the usual area frame method of delineating frame units based on predetermined and more loosely defined strata.(3)(5)

Stratification technicians indicated that the DBS technique was less difficult, primarily because the time consuming and difficult search for city stratum boundaries was eliminated. As the next section indicates, the costs were also much lower.

Costs

In response to several questions generated by the first DBS report (1), a detailed cost breakdown for sampling and enumeration is presented.

Sampling Costs

The hours spent by cartographic personnel on frame construction and sampling were as follows:

Task	Hours
Stratification	2047
Digitization	518
Sample selection	243
Pen and ink	387
TOTAL	3195

This is an average of 15.6 hours per segment, which includes defining auxiliary data for the frame units, compared to 20 or more hours per segment for the JES in a state with section lines. To obtain a cost which we can add to the enumeration expenditures, these hours must be amortized over the life of the frame and rotation sampling costs must be considered.

Assume that the frame has a 15 year life and that the 205 segments have a 5 year rotation cycle. Frame construction hours can be associated with each year that a segment is used by first adding the hours spent on stratification and digitization and then dividing by the product of the number of segments used in a year and the number of rotation cycles in the frame.

This computation yields:

$$\frac{2047 + 518}{(205)(15/5)} = 4.2 \text{ hours per segment selected}$$

Each segment is used for 5 years, so the construction cost is 4.2/5=.9 hours per segment-use year. Sample selection costs are computed by adding the hours spent on sample selection and pen and inking and then dividing by the number of segments selected the first year. This computation yields:

$$\frac{243 + 387}{205} = 3.0 \text{ hours per segment,}$$

or 3/5 = .6 hours per segment-use year. Adding the two portions results in 1.5 hours per segment use year as the sampling cost under the above assumptions. Frame materials, excluding LANDSAT scenes, cost about $600.00.

The cost analysis outlined above is based on certain assumptions and approximations. What is important in the analysis is not necessarily the "to-the-dollar" accuracy of the costs, but the effect these assumptions have on the conclusions. In this analysis, changes in the life of the frame and/or the rotation cycle would certainly affect the cost estimates presented above.

The assumption of a 15 year life of the frame is based on experience of other midwestern states' area frames. The way the frame was constructed (small count units which can be easily updated) also lends itself to a longer life since minor restratifications can be done with little manual input. It is reasonable, then, to assume a useful frame life of at least 15 years.

If the five year rotation cycle were different, the construction cost per segment would vary but the cost per segment use year would remain constant. Thus, the sampling cost of 1.5 hours per segment-use year is also reasonable.

Enumeration Costs

The questionnaire used in the DBS was similar to that used in the JES for collecting data on crop acreages in the segment. It was designed to obtain data on all varieties of dry beans and all other crops. Data was not collected for livestock or economic items. In 1981 both tract data and entire farm data were collected, in order to compute a weighted estimate. In 1982, entire farm data were not collected.

Twenty-one enumerators attended a one-day training school held in early July each year. The enumeration period was from approximately July 7 through July 22. Whenever possible, data were collected by personal interview with the farm operator. Data collection by observation was used as a last resort.

The reduced cost in 1982 (Table 3) can be attributed to the lack of rotation, enumerator experience with the survey and, primarily, not collecting entire farm data. Considering frame development and sampling at $6.00 per hour and professionals salaries as a fixed cost not to be included, annual survey cost is:

$$\$6.00 \times 1.5 \times 205 + \$13{,}000 + \$600 \; / \; 15 = \$14{,}576.$$

Comparison with the JES

Table 4 compares estimates and their precision between the 1982 JES and DBS. The 16 county area of the DBS contains a large portion of the land planted to Michigan's major crops. The statistics for dry beans support the hypothesis which led to the DBS. That is, the JES collects data before most of the dry bean crop is planted and, therefore, cannot accurately estimate planted acres. The shortcoming of the JES is the result of economic and natural conditions. Dry beans can be planted later than many other crops, and are therefore likely to be involved in changes to earlier intentions. Weather and economic conditions may cause growers to plant dry beans instead of some other crops that may have been intended to be planted at the time the JES is conducted. These circumstances have resulted in a historic underestimate of dry bean acres from the JES.

Table 3 - Actual enumeration costs reported by the Michigan SSO

	1981 Costs	1982 Costs
Regular Salary	$11,238	$7,163
Overtime	722	1,473
Fringe Benefits	4,262	574
Mileage	3,823	3,400
Payroll Costs	500	--
Telephone	108	71
TOTAL	$20,653	$12,681

Table 4-Comparison of estimates from the 1982 June Enumerative and Dry Bean Surveys

	1982 JES 1/			1982 DBS 2/		
Crop Planted	Direct Expansion Estimate	Standard Error	Coefficient of Variation	Direct Expansion Estimate	Standard Error	Coefficient of Variation
	(1,000 acres)		Percent	(1,000 acres)		Percent
Corn	3,217.2	156.4	4.9	1,282.2	77.5	6.0
Soybeans	1,041.6	92.8	8.9	606.8	57.9	9.5
Sugar Beets	0.0	0.0	0.0	76.5	16.1	21.0
Oats	526.8	51.4	9.8	233.1	21.0	9.0
Winter Wheat	557.8	63.4	11.4	264.0	27.7	10.5
Dry Beans	382.9	56.6	14.8	522.4	41.2	7.8

1/ Entire state
2/ 16 counties

Since the DBS is conducted only in the 16 county area the estimate of dry beans is expected to be slightly lower than the state total for dry beans. Thus, a one tailed test of the hypothesis that the DBS estimate is equal to the JES estimate versus the alternative that the DBS estimate is larger was in order. The hypothesis of equality is rejected at about the 5 percent level.

Estimates for Other Crops

The DBS level of precision for the estimates of other crops is of interest. Consider a combined post-stratified (multiple frame) estimator of the following type:

$$\hat{X} = X_{JES,OUT} + (A\,X_{JES,IN} + B\,X_{DBS})$$

where $A + B = 1$

$X_{JES,OUT}$ = The JES domain estimate for the counties not included in the DBS.

$X_{JES,IN}$ = The JES domain estimate for the counties included in the DBS.

X_{DBS} = The DBS estimate.

Then $\hat{X}$ is an unbiased estimate (provided that the subscripted X's are unbiased) with variance:

$$V(\hat{X}) = V(X_{JES,OUT}) + A^2 V(X_{JES,IN}) + B^2 V(X_{DBS}).$$

Based on the estimates from Table 4, assume that the poststratified variance for a selected crop from the JES is proportional to the ratio of the estimated acres in the poststratified area to the estimate for the state. Choosing B=1 would result in X having about an 8 percent reduction in standard error for corn over the JES estimate and about a 15 percent reduction for oats. Since the computation of these estimates is simple, the SSO can use the DBS to obtain late season estimates with less bias or lower variance for other crop acreages at no additional out of pocket expense simply by choosing the appropriate A and B for each crop.

Other Survey Designs

In very few applications can the statistician be certain that the "best" estimator is being used when cost, precision, and accuracy are the criteria. Given a particular sampling frame and survey design, optimum allocations can be developed. Yet this is optimum only for the design and frame in use. There is rarely a way to determine how much better an optimum allocation would be for a different design. It is still harder to determine how much better a different frame would interact with various designs and their optimum allocations. In an ideal situation, the mathematical statistics concept of admissability can be formulated for the choice of sampling frame and design. In most applications, sufficient data do not exist to formulate admissability rules exactly, and experienced judgment criteria is used to choose the foundation under which the estimators are optimized. This section will discuss various other frames and designs which could have been used along with their perceived strengths and weaknesses.

The JES Frame

The JES area frame could have been used for the DBS, but analysis of the variances from the 1980 JES indicated that with a sampling rate equivalent to the DBS, the best that frame could do would be to meet the maximum variance desired from the special survey (one with an 8 percent CV). Also, using the JES frame would have precluded some research on frame construction. Specifically, research concerning the use of control data to create strata after frame units are defined and the use of soil type information would have been restricted. This section will outline the expected costs and precision had the JES frame been used for the DBS.

As mentioned earlier, data collection in May does not meet the accuracy requirements of dry bean acreage estimation. The timing of data collection needed to be in July and returning to the same JES segments later was ruled out as a respondent burden concern. Therefore, the JES sample could not be used, and a new sample would have to be drawn.

Stratification and digitization costs could be reduced to a record keeping chore by using selected paper strata from the JES frame. This task would entail about 200 hours. Frame materials would cost the same amount as the DBS. Sample selection would take longer using the JES frame because there would be more frame unit splits, more half section lines used, and selections made in the more difficult city and range strata. Selection time would be about 1.7 hours per segment. Pen and Ink would take about 2 hours per segment. Using the same cost evaluations of the previous section, the development and sampling costs would be slightly over .8 hours per segment-use year compared to 1.5 hours for the DBS frame, for a saving of about $4.00 per segment per year.

Since segments in the JES are one half square mile in size, while the DBS segments are one square mile, enumeration costs would be lower, but certainly not by a factor of one half. The average farm size in Michigan was estimated at 177 acres in 1982 (6). The 1978 Census of Agriculture indicates that the average size is over 200 acres in the dry bean area. These data indicate that a half mile segment in this area would contain somewhat more than half the agricultural tracts of a one mile segment which includs the half mile area. Adequate statistics were not readily available to quantify the exact difference, but about a one-quarter to one-third reduction seems reasonable. This would mean a total survey cost of between 4 and 5 thousand dollars less. Based on the JES estimates of variance, using this money for additional segments would result in variances similar to the DBS. Since the JES acreage estimates have been biased low, experience indicates that the estimated variance would also be low. Thus at the same cost the JES frame does not seem able to achieve the precision of the DBS for an unbiased estimate.

List Frame

As indicated in the first report, efforts to improve the list and multiple frame surveys were not successful. The cost and time required to maintain a list was considered undesirable by the SSO. The large resource requirements were necessary because there are large annual changes in the number of producers and the quantities they produced. For the past several years the list frame has not been used. It would be worth examining the benefits of a large operator list frame with a multiple frame estimator from the DBS.

EXPECTATIONS FOR THE 1983 SURVEY

A minor reallocation will be made for the 1983 DBS. Twenty-one segments, 10 percent of the total, are to be reallocated. This reallocation is expected to reduce the CV from eight percent to seven percent. Changes in location or size of fields could alter this expected improvement, but this is not anticipated since the relationships did not change while the planting decreased between the two years.

An annual rotation of segments is not considered necessary at this time because growers are very receptive to the survey. The high DBS response rate reported by the SSO coincides with indications in a study in the Dakotas (2) that response rates are related to the respondents' perception of the relevance of the survey material.

The regression estimators may be computed, but no additional gains are expected as no changes are to be made to the auxiliary data. If a ratio estimator is desired, a matched sample estimator will need to be computed. Full state estimates might best come from the combined area frame domain approach described earlier. Starting in 1983, summarization will be done by the SSO with first-year assistance by the Sampling Frame Development Section.

PROPOSALS FOR FUTURE RESEARCH

The DBS presents a unique and relatively inexpensive research opportunity for assessing the potential of various survey designs and estimation techniques. Combined and multiple frame estimates have been suggested earlier in this paper as topics for future research. This section presents several other ideas for future research utilizing the DBS. The purpose here is not to present fully developed plans, but to initiate discussions which could lead to project proposals.

Costs

The cost associated with using a particular survey is as important in choosing the design as the precision and accuracy. Area frame surveys in general have little cost information available. Even the totals for enumeration, salary, mileage, development cost, etc. (which are available in SSO's) are not well documented. In many cases, other costs which are important are not available.

For example, consider the decision to construct a new frame for a state. The high initial costs of constructing a new frame can be amortized over the life of the frame. The question to ask is whether it is cheaper to add segments to the old frame, to update only portions of the old frame or to construct an entirely new frame. The answer lies in the marginal cost of additional segments, a cost about which little is known. The replicated design can be used to estimate these costs by adding replications to a fixed frame over several years. The replicated design can also be used to test for optimal segment size via a split sample approach.

Soils

Soil type was used in stratification and estimation in an effort to improve the precision of area frame estimates. In the dry bean survey, some increased precision is evident for dry beans as well as the other crops, but it was not as much as hoped. It is important not to interpret this as cause to end research in the use of soils. It is obvious from the DBS that soils can be useful, athough the way soils were quantified was not as effective as hoped. The soils maps delineated broad soil groupings. Within each grouping, there are usually four to six sub-soil types. The broad soil grouping were not based on agricultural potential and thus often were composed of a mixture which ranged from soils suitable for dry bean planting to those not at all suitable.

Soil scientists at Michigan State University computed the potential yield per acre for dry beans for each sub-soil type and a weighted average for the soil groups, using the percent of area having the soil as the weight. In the 1981 DBS this weighted average potential yield was the numeric quantity used for stratification and regression estimation. As a result, many soil groups had moderately good ratings, but were quite variable in the potential for dry beans within the group. A review of the segment data showed that dry beans did not appear on the poor soil types, while these variable soil groups were a major source of within strata variance. This experience reinforces two statistical lessons: (1) the average is not always the best statistic and (2) the auxiliary variable should be as highly correlated with the item of interest as possible.

Soils were found to be associated with crops. What is needed is a quantification of the soils which is statistically correlated with the plantings. In 1982 the best soil type in the group was used as the quantifier in the regression estimates. This too was not a highly satisfactory quantifier.

The question is then how soils can be quantified so that they can be useful. The answer lies not in trying to create a numerical value for the soils groups which were used, but in creating the proper soil groups. Trying to force standard soil grouping to help in an application for which they were not intended is not satisfactory. To effectively use soils, groupings must be made for the specified application. This is the direction in which research should continue. Identification of either more detailed soil reference materials or a soils research group that can make the necessary analysis is the first step. This is the most formidable task since it will take exploration outside SRS in a field other than statistics. It seems that this task can produce good results in area and possibily yield forecasting which would be worth the work and persistence needed.

Stratification

The soil variable used to stratify resulted in some highly variable strata. As research dictates changes to the frame, these strata should be identified via ANOVA techniques and should be the first to be collapsed and reworked using more statistically efficient techniques.

REFERENCES

(1) Fecso, Ron, Jeff Geuder, Bob Hale and Steve Pavlasek, Estimating Dry Bean Acreage in Michigan , U.S. Department of Agriculture, Statistical Reporting Service, Staff Report No. AGES820225, 1982.

(2) Fecso, Ron and Robert D. Tortora, Farmers Additudes Toward Crop and Livestock Surveys: A colletion of Papers Related to the Analysis of the Survey of Dakota Farmers and Ranchers , U.S. Department of Agriculture, Statistical Reporting Service, Staff Report No. AGES811007, Nov. 1981.

(3) Fecso, Ron and Van Johnson, The New California Area Frame: A Statistical Study , U.S. Department of Agriculture, Statistical Reporting Service, No. 22, 1981.

(4) Fecso, Ron, Van Johnson and Jeff Geuder, Using SAS to Evaluate an Area Sampling Frame For Agricultural Surveys, Proceedings of the Sixth Annual SAS Users Group International Conference, 363-368, Orlando, Fl., Feb. 1981.

(5) Houseman, Earl, Area Frame Sampling in Agriculture, U.S. Department of Agriculture, Statistical Reporting Service, No. 21, 1975.

(6) Michigan Department of Agriculture, Michigan Agricultural Statistics , Michigan Agricultural Reporting Service, Lansing, Michigan, Aug. 1982.