METHODOLOGY TO EVALUATE SAMPLE DESIGNS OF THE QUARTERLY AGRICULTURAL SURVEY

Robert J. Battaglia, National Agricultural Statistics Service Room 204 Health and Agriculture Building, CN-330 New Warren Street, Trenton, NJ 08625

KEY WORDS: Multivariate optimal allocation, sample design, quarterly agricultural surveys.

ABSTRACT, The sample design of the Quarterly Agricultural Survey was evaluated by comparing allocations and precision measures for the operational program vs. a multivariate optimal allocation. The optimal list sample allocations computed for each state were based on major crop and livestock items and grain storage capacity. Results showed the optimal allocations to be useful in determining the sample size necessary to obtain target levels of precision. The optimal allocations will also be useful when planning allocation of samples to strata and the removal of ineffective strata.

SUMMARY, In October, 1986 NASS created the Sample Design Section (SDS) as part of the reorganization of Headquarters. One of the major duties of SDS is to evaluate the sample sizes of the probability surveys. The purpose of this report is to demonstrate the procedures that SDS will use to evaluate the list sample sizes of the Quarterly Agricultural Surveys (QAS).

Any evaluation of sample sizes focuses on the coefficients of variation (CV's). This report compares the CV's that NASS obtains operationally vs. the CV's that NASS desires. This report makes these comparisons at the state level because the state level is: 1) where sample sizes are established for QAS and 2) where most of the differences occur between operational and target CV's. When there is a discrepancy between operational and target CV's, then three components must be evaluated that affect the CV's of each multiple frame indication -- the list CV, the nonoverlap CV, and the percent of the indication which comes from the nonoverlap (NOL) domain. The last component directly involves list frame coverage, which affects any evaluation of sample sizes.

This report has five tables for each state. The first table compares operational vs. target levels for the CV of the multiple frame indication, the CV of the list indication, the CV of the NOL indication, and the percent of the indication that came from the NOL. The target for each component shows how much that component would have to change if all other components were left at their operational levels. The second table shows the change in list sample sizes that will result from using optimum allocations in each state to achieve the target CV's. The third table gives the list CV's that will result for each indication if NASS uses the optimum allocation. The fourth table shows the operational vs. optimum allocations for the strata of the hogs/crops sample, and the fifth table makes this comparison for the cattle/sheep sample. With each set of the tables, there are comments for each state.

An important factor in this report is the level of each target CV. This report uses target CV's that SDS established to conform with PSM45-88 and to reflect the size of each state's indication in relation to other states. These target CV's are not "engraved in stone", but an effort has been made to set them at practical levels in order to demonstrate the procedures. These targets should be reviewed before changes in QAS sample sizes are formally recommended in the future. The levels of the targets reflect national needs and not the needs that individual states may have for the indications.

Few changes in the operational program will be made from this report -- its purpose is mainly educational. However, in the future SDS will suggest changes to the operational program after the procedures have become familiar and more QAS data has been obtained for 1987 and 1988. Of course, any formal changes in sample sizes require specification and budget approval.

METHODOLOGY TO EVALUATE SAMPLE DESIGN OF THE QUARTERLY AGRICULTURAL SURVEY

TECHNICAL BACKGROUND, The QAS survey provides state, regional and national indications of acreage and production for major crops, livestock inventories, and on-farm grain stocks [4]. Two sampling frames, area and list, together provide complete coverage of all farm operations. The area portion of the QAS sample is a subsample of nonoverlap tracts from the June Enumerative Survey while the list portion of the sample comes from the List Sampling Frame (LSF) maintained by each NASS field office.

Prior to the QAS, separate surveys were conducted for cattle, hogs, crops, sheep and grain stocks -- with each survey having its own sample design. Strata allocations for these surveys were usually determined by using optimum allocations on the one major item from the survey. The optimum allocation minimizes the CV of an indication [3].

With the Integrated Survey Program (ISP), surveys for hogs, crops, cattle, sheep and grain stocks were combined into one survey. Although approaches to the problem of multivariate allocations were known (Huddleston, Claypool, and Hocking [4]), they were not used operationally. Univariate optimum allocations were computed for major items of interest and the strata allocations were subjectively determined from the univariate results.

Bethel [1] developed an algorithm to perform multivariate optimum allocations. NASS used this algorithm, which can be run on minicomputers, to determine optimum allocations for the 1985 ISP/JES stratification [2]. These optimums were never used however, since program changes resulted in two surveys, one for hogs/crops and one for cattle/sheep--the present QAS design.

The purpose of this research was to evaluate the list sample sizes of QAS. Multivariate optimal allocations were compared to operational allocations. Sample sizes could then be determined that will produce desired coefficients of variation (CV's) for several commodities. Also, some ineffective strata could be identified.

PROCEDURES, Optimum Allocations

The multivariate optimal allocation algorithm was used to produce QAS Allocations for the non-preselect list strata based on the 1986 December QAS and the 1987 January QAS survey data.

Inputs for the optimal allocation algorithm were obtained from sample allocation worksheets. These worksheets are produced after a QAS summary and contain the information required for the optimum allocation by state for the major QAS items. Target MF CV's were determined considering recommendations from PSM45-88, the importance of the item relative to the U.S. level, and the sample size necessary to obtain the target. Target list CV's were then calculated based on these target MF CV's and the above mentioned survey data.

Of the commodities used for the optimum allocation, cropland acres and total hogs were used for each state. Grain storage capacity was used in each state except for the six New England states where hay stocks was substituted. The remaining commodities were different from state to state. For example, potato acreage was selected if the state was part of the potato objective yield program and the sample allocations contained potato strata. Generally, specialty crops were excluded except for peanuts and tobacco in a few states where the acreage was significant. Since the program which computes the optimal allocation algorithm is limited to ten variables, no state could have more than 10 items considered from the December QAS.

Inputs for the January QAS were also obtained from sample allocation worksheets. Total cattle and total milk cows were used in all states while total sheep was used in states with significant sheep numbers.

Design Components, The CV for a multiple frame indication is affected by three components -- the CV from the list portion of the sample, the CV from the nonoverlap (NOL) portion of the sample, and the percentage of the indication from the NOL domain. This relationship is shown in the following formula:

$$(\text{MF CV})^2 = (\text{list CV})^2 * (1-p)^2 + (\text{NOL CV})^2 * p^2$$

where:

MF CV = multiple frame coefficient of variation.
list CV = coefficient of variation from list domain.
NOL CV = coefficient of variation from NOL domain.
p = proportion of indication from NOL domain.

The above relationship can be algebraically manipulated to solve for the level of a component needed to reach a target MF CV level.

RESULTS, Results for each state are in the Appendix. Five tables were produced for each state to evaluate the QAS sample sizes. After describing each type of table, the tables for Alabama will serve as an example to describe the interpretation of the results.

Comparison of Factors Affecting QAS Sample Design, Table 1 gives a comparison of multiple frame CV's, list CV's, nonoverlap CV's and percentages of the indications coming from the nonoverlap. For each of these factors, the value that was obtained operationally is compared to a target value. The target for each multiple frame CV was set by SDS and conforms with PSM45-88. The multiple frame target also reflects the size of each state's indication relative to the U.S. indication. The targets for the other components reflect how much each component would have to change by itself in order to achieve the target for the multiple frame CV.

For Alabama, the target for the multiple frame CV of cattle was 5.0. However, the operational CV in January 1987 was 4.8 -- below the target. Thus, if any one component -- list CV, nonoverlap CV, or percentage nonoverlap-- increased a small amount, the target for the multiple frame CV would still be achieved. All three components could not increase at the same time, but any one component could increase. All of the commodities in Alabama were at or below the target for the multiple frame CV except for hogs and cotton.

Cotton in Alabama has a target for the multiple frame CV of 15.0. the multiple frame CV from the December 1986 QAS was 15.4. To reach the target MF CV, one of the components must be reduced. Either:

1) the list CV must be lowered to 11.2, or 2) the nonoverlap CV must be lowered to 26.8, or 3) the percent nonoverlap must be lowered to 47.6. Any one of these operations would lower the multiple frame CV to 15.0. The multiple frame CV for cotton could also be reached by lowering a combination of the three components. While the best combination of components can not be determined algebraically, these tables will be useful for determining which component can most easily be reduced to reach the target level of precision.

The entries for total hogs in Alabama will also be examined. The target for the multiple frame was 10.0 while the multiple frame CV from December 1986 was 24.1. Table 1 shows that no single component can be reduced enough to cause the multiple frame CV to decrease to 10.0.

In this situation the target CV of 10.0 is so much smaller than the operational CV of 24.0 that it is impossible to achieve the target by varying any one component. NASS could raise the target CV to a more achievable level, but that action would disguise what is basically a deficiency in the current precision of the QAS. The better options might be: 1) to wait and look at results from other quarters of the QAS since outliers may be affecting these results or 2) to try to reduce all of the components.

Effect of Target CV's On List Sample Sizes, Table 2 shows the effect on list sample sizes in Alabama when the target list CV's are put into a multivariate optimal allocation. Separate allocations are given for the hogs/crops survey and the cattle/sheep survey. The table shows that a sample size of 1253 from the list strata will achieve the target list CV's for the hogs/crops survey. The sample size is 123 larger than the operational sample size used in December 1986. (Total hogs had no effect on the overall sample size since it was impossible to achieve the target of the multiple frame CV by only changing the list sample size.) The optimal allocation for the cattle/sheep strata recommended a sample size of 570 to obtain the targets for the list CV's. This sample size was a decrease of 336 from the sample size used in the January 1987 survey. The decrease in the cattle/sheep sample size resulted because the targets of the list CV's in Table 1 were higher than the operational CV's for total cattle and total dairy cows.

Very large increases in the list sample size are required for some states when the target list CV is very low. In these situations, the analysis indicates that the target list CV and the target MF CV can be reached with a large increase in the list sample size. Usually the cost would be prohibitive so steps to achieve the target NOL CV or the percent NOL should be examined. Another solution would be to increase the target MF CV level.

Comparison of Target List CV's and the List CV's Obtained Under the Optimum Allocation, The third table compares the target list CV's from Table 1 to the list CV's that would occur with the multivariate optimal allocation. Since a target list CV was for hogs, the operational list CV (25.8) was substituted into the optimal allocation. The list CV for corn under the optimum allocation (7.0) is less than the target List CV (9.4). This result shows that the sample size needed to achieve the target for the other variables has actually helped the corn CV to go below its target.

Comparison of Operational and Multivariate Optimum Allocations for Strata 60-94 of the December 1986 QAS, The fourth table compares the sample sizes by strata for the hogs/crop survey. The operational sample size from the December 1986 QAS is shown with the sample size from the multivariate optimum allocation. These tables can be used to identify ineffective strata and to change strata sample sizes.

For the hog/crop strata in Alabama, Table 4 shows that the optimal allocations for strata 62, 64, and 66 were smaller than the operational allocation while allocations for strata 63, 67, and 70 should be increased. The optimal allocation for stratum 82 (sweet potatoes) showed a decrease because this item was not included in the optimal allocation. Also, the effect of this stratum on the items used in the optimum allocation was not large.

Comparison of Operational and Multivariate Optimum Allocations For Strata 1-34 of the January 1987 QAS, Table 5 is in the same format as Table 4, but Table 5 contains results for the cattle/sheep survey. Table 5 shows the optimal allocation to be less than the operational allocation for all strata because the target list CV's were greater than the survey list CV's for both items. Strata 1, 3, and 4 showed significant decreases from the operational allocation.

RECOMMENDATIONS, The purpose of this research was to demonstrate an evaluation of the sample sizes for the Quarterly Agricultural Survey (QAS). Operational allocations were compared to multivariate optimum allocations for the list strata of the hogs/crops sample and the cattle/sheep sample.

Based on the findings, we recommend:

1. The targets for the multiple frame CV's be carefully reviewed for each state so that future evaluations of the QAS

will have accepted targets as a basis for analysis.

2. The procedures in this report be considered as part of an acceptable process for evaluating the sample sizes of probability surveys. At least two years (and preferably three years) of QAS data will be analyzed before major changes in QAS sample sizes are formally submitted for specification and budget approval.

REFERENCES,

1. Bethel, James W. (1985). Sample Design for the 1985 ISP/JES. Staff Report No. 86. U.S. Dept. of Agriculture, Statistical Reporting Service.
2. Bethel, James W. (1986). An Optimum Allocation Algorithm for Multivariate Surveys, Staff Report No. 89. U.S. Dept. of Agriculture, Statistical Reporting Service.
3. Cochran, William G. (1977). Sampling Techniques, 3rd Edition, John Wiley & Sons.
4. Huddleston, H.F., P.L. Claypool, and R.R. Hocking. (1970). "Optimum Sample Allocation to Strata Using Convex Programming," Applied Statistics, Vol. 19, No. 3, pp.273-278.
5. U.S. Department of Agriculture, National Agricultural Statistics Service. (1986). December Supervising and Editing Manual.

State: ALABAMA

Table 1. Comparison of factors affecting QAS sample design

		Multiple Frame CV	Components of Multiple Frame CV: List CV	Percentage Nonoverlap	Nonoverlap CV
CATTLE	Jan 87	4.8	2.8	43.6	10.4
	Target	5.0	3.7	45.8	10.9
DAIRY	Jan 87	9.3	5.3	10.4	77.4
	Target	10.0	6.6	11.4	84.6
CORN	Dec 86	10.0	9.4	36.3	22.0
	Target	10.0	9.4	36.5	22.1
SOYBEAN	Dec 86	9.8	9.3	46.5	18.1
	Target	10.0	10.1	48.5	18.7
WHEAT	Dec 86	14.7	12.8	40.5	30.9
	Target	15.0	13.9	42.2	31.9
CROPLAND	Dec 86	9.9	4.9	55.2	17.5
	Target	10.0	5.8	55.8	17.7
HOGS	Dec 86	24.1	25.8	34.1	43.1
	Target	10.0	*	*	*
CAPACITY	Dec 86	9.9	5.8	34.0	27.0
	Target	10.0	6.0	34.2	27.2
COTTON	Dec 86	15.4	13.5	50.1	27.8
	Target	15.0	11.2	47.6	26.8

* Target was impossible to establish since the target for the multiple frame CV was so low.

Table 2. Effect of target CV's on list sample sizes for strata that were not prob 1's.

Dec 86	Crop/Hog	Cattle/Sheep
Optimum Allocation To Achieve Target List CVs	1253	570
Change From Operational QAS	123	-336

Table 3. A comparison of target list CV's and the list CV's obtained under the optimum allocation.

	Target List CV	List CV Under Optimum Allocation
CATTLE	3.7	3.7
DAIRY	6.6	6.6
CORN	9.4	7.0
SOYBEAN	10.1	8.7
WHEAT	13.9	9.2
CROPLAND	5.8	5.8
HOGS	25.8*	25.0
CAPACITY	6.0	6.0
COTTON	11.2	10.1

* Target list CV was impossible to establish since the target for the multiple frame CV was so low. The operational list CV was substituted in the optimal allocation.

Table 4: A Comparison of Operational and Multivariate Optimum Allocations for Strata 60-94, of the December 1986 QAS.

SURVEY: HOGS/CROPS

QAS STRATA	LSF STRATUM DESCRIPTION	N_h	DEC 86 SIZE	OPT ALLOC
60	ALL LAND 100-999	4162	75	54
62	CROPLAND 1-29	4448	130	26
63	CROPLAND 30-399	4736	190	290
64	CAPACITY 1-4999	1643	85	36
66	HOGS 1-49	2415	160	20
67	CROPLAND 400-3999	1082	110	210
68	ALL LAND 1000+	218	25	6
70	CAPACITY 5K-74999	944	115	487
75	HOGS 50-99	246	35	39
80	HOGS 100-199	209	50	32
81	HOGS 200-499	191	65	26
82	SWEET POTATOES	136	55	6
83	HOGS 500-1499	70	35	21
TOTAL		20500	1130	1253

Table 5: A Comparison of Operational and Multivariate Optimum Allocations for Strata 1-34 of the January 1987 QAS.

SURVEY: CATTLE/SHEEP

QAS STRATA	LSF STRATUM DESCRIPTION	Nh	JAN 87 SAMPLE	OPT ALLOC
1	1-49 CATTLE	14167	350	196
	50-99 CATTLE	3076	140	116
3	100-199 CATTLE	1611	135	82
4	200-499 CATTLE	789	130	60
10	50-99 DAIRY	107	18	16
11	100-199 DAIRY	110	25	23
20	500-999 CATTLE	175	60	45
21	1000-1999 CATTLE	45	28	18
24	200-299 DAIRY	30	20	14
TOTAL		20110	906	570