United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Survey Sampling
Branch

NASS Staff Report
Number SSB-90-01

April 1990

# EFFECTS OF DESIGN DEVIANT ALLOCATION

James W. Mergerson

# Effects of Design Deviant Allocation.

James W. Mergerson. Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, April 1990, Staff Report Number SSB9001.

## ABSTRACT

Effect of deviations from a design based allocation for an area frame sample is investigated. The effect of the deviations was small.

**Keywords:** stratified sampling, optimal allocation, area frame

**This paper was prepared for internal distribution to Washington, D.C. Staff**

### Table of Contents


| | Page |
|---|---|
| SUMMARY | i |
| INTRODUCTION | 1 |
| ANALYSIS | 2 |
| CONCLUSIONS | 7 |
| REFERENCES | 7 |

## SUMMARY

The National Agricultural Statistics Service (NASS) uses an area sampling frame with land-use stratification to obtain farm related estimates. Prior to use of a new frame, sample sizes by substrata must be determined. A design based sample allocation to substrata is not used. Sample units are allocated to strata and partitioned equally among substrata within each stratum. The strata-level allocation results in a slightly inflated State sample size. Overall, having equal substrata sample sizes within each strata instead of allowing the substrata sample sizes to vary showed insignificant effect for Kansas in this study.

# Effects of Design Deviant Allocation

James W. Mergerson
USDA-NASS-RAD-SSB
Area Frame Section

## INTRODUCTION

The purpose of this analysis is to evaluate the effect of computing new area sampling frame sample allocations using stratum level estimated standard deviations instead of estimated standard deviations at the sub-strata level. Potential gains in reallocating based on the sub-stratified design and allowing sample size to vary among sub-strata are also investigated.

The National Agricultural Statistics Service (NASS) uses area sampling frames in conducting surveys to obtain information regarding crop acreages, cost of production, farm expenditures, livestock inventories and other items. There is an area frame for each state except Alaska. NASS area frames are partitions of total land area within States according to land use. Land-use strata are further partitioned into geographic areas of similar agriculture based on county level data. Normally, the Area Frame Section of NASS constructs new area sampling frames for two States each year. After the frame is constructed and prior to its use, sample units must be allocated.

This initial allocation is performed at the stratum level. Once the stratum level sample sizes are determined, most strata are further partitioned into sub-strata and sample units are divided equally among assigned sub-strata within a stratum. A replicated sample is selected in each stratum. The exact number of sub-strata within a stratum is determined jointly with the number of replicates after computing the optimum stratum level sample size. Thus, design based allocations are not currently done.

## ANALYSIS

Analyses were performed using 1987 and 1986 Kansas June Agricultural Survey (JAS) data. Area frame design information for the Kansas frame is given in table 1.

Table 1. Kansas Area Sampling Frame Design Information - (1975-1989).

| Stratum | Number of Segments In Population | Target Segment Size | Number of Segments In Sample | Number of Sub-strata | Stratum Definition |
|---|---|---|---|---|---|
| 11 | 25,070 | 1.00 | 170 | 17 | More than 75% cultivated |
| 12 | 21,736 | 1.00 | 120 | 12 | 50-75% cultivated |
| 20 | 21,284 | 1.00 | 100 | 10 | 15-49% cultivated |
| 31 | 2,845 | 0.25 | 12 | 3 | Agri-Urban |
| 32 | 2,941 | 0.10 | 12 | 3 | Residential-Commercial |
| 33 | 246 | 0.25 | 2 | 1 | Resort |
| 40 | 3,163 | 4.00 | 15 | 3 | Range |
| 50 | 293 | 1.00 | 2 | 1 | Non-Agricultural |
| 61 | 29 | 0.50 | 2 | 1 | Potential water |
| Total | Sample | Size | 435 | | |

The analysis included total cattle and calves and planted acreages of corn, soybeans, sorghum and winter wheat. These items were selected based on their importance to the State's agricultural statistics estimating program.

Allocation computations are obtained using a multivariate optimal allocation program. The original version of the program was developed by Bethel (1986). Mergerson (1986) and others modified the original program and implemented it in various computing environments for operational use. The use of the program on a personal computer is described by Mergerson (1988). Some of the required inputs are counts of the number of possible sampling units in each stratum or sub-stratum; state level estimates of population totals and target coefficients of variation (CVs); and strata or sub-strata estimates of standard deviations. Both 1987 and 1986 estimates and CVs are shown in table 2. One set of CVs (strata) are computed ignoring substratification. The other set of CVs (sub-strata) are computed relative to the current design.

These sets of input CVs provide a comparable basis for comparing strata based allocations to design based allocations.

Table 2. Kansas strata versus sub-strata allocation analysis
Input State estimates and coefficients of variation

| | Survey Estimates (000) | | Coefficients of Variation Strata | | Sub-strata | |
|---|---|---|---|---|---|---|
| Item | 1987 | 1986 | 1987 | 1986 | 1987 | 1986 |
| Cattle | 1,600 | 1,450 | 13.7 | 15.7 | 13.7 | 15.6 |
| Corn | 1,300 | 1,500 | 14.1 | 14.0 | 13.7 | 14.0 |
| Soybeans | 2,300 | 2,000 | 10.3 | 10.0 | 8.2 | 7.7 |
| Sorghum | 3,800 | 4,300 | 6.0 | 5.6 | 5.9 | 5.5 |
| Wheat | 10,750 | 11,400 | 3.9 | 4.1 | 3.2 | 3.5 |

Table 2 estimates and CVs were used along with estimated standard deviations to compute optimum allocations. Optimum allocations are rounded to a multiple of five to facilitate a five year rotation cycle. That is, about 20% of the sample units are replaced each year with new sample units. Strata based allocations are rounded at the strata level. Sub-strata based allocations are rounded at the sub-strata level. Tables 3 and 4 list optimal and rounded allocations summarized at the stratum level based on both strata and sub-strata level input standard deviations for the years 1987 and 1986 respectively. Strata 31, 32, 33, 50 and 61 are excluded from the tables since the optimum allocation to these strata is at most five sample units.

Table 3. Kansas strata versus sub-strata allocation analysis
1987 Strata-level sample size comparisons

| Stratum | Current | Optimal Strata Based | Optimal Sub-strata Based | Rounded Strata Based | Rounded Sub-strata Based |
|---|---|---|---|---|---|
| 11 | 170 | 166 | 151 | 165 | 150 |
| 12 | 120 | 122 | 112 | 120 | 115 |
| 20 | 100 | 95 | 94 | 95 | 95 |
| 40 | 15 | 20 | 20 | 20 | 20 |
| Sub-Totals | 405 | 403 | 377 | 400 | 380 |

Table 4. Kansas strata versus sub-strata allocation analysis
1986 Strata-level sample size comparisons

| Stratum | Current | Optimal Strata Based | Optimal Sub-strata Based | Rounded Strata Based | Rounded Sub-strata Based |
|---|---|---|---|---|---|
| 11 | 170 | 166 | 149 | 165 | 155 |
| 12 | 120 | 124 | 111 | 125 | 110 |
| 20 | 100 | 90 | 88 | 90 | 90 |
| 40 | 15 | 22 | 20 | 20 | 25 |
| Sub-Total | 405 | 402 | 368 | 400 | 380 |

The rounded strata based allocations were 20 sample units greater than the sub-strata based allocations for both years. Thus, the operational sample allocation is slightly larger than it would be if it were design-based.

The effect of allowing the sub-strata sample sizes to vary instead of keeping them the same size within a stratum can be evaluated by comparing expected CVs. Expected CVs relative to the current frame are compared to expected CVs computed using rounded unequal substrata sample sizes. Expected CVs are shown in table 5.

Table 5. Expected coefficients of variation (CVs) using equal sub-strata sample size within strata versus using unequal sub-strata sample size within strata (both sample sizes are multiples of five).

| Item | 1987 Expected CVs Equal Sub-strata | 1987 Expected CVs Unequal Sub-strata | 1986 Expected CVs Equal Sub-strata | 1986 Expected CVs Unequal Sub-strata |
|---|---|---|---|---|
| Cattle | 13.5 | 13.6 | 15.5 | 13.9 |
| Corn | 13.6 | 13.4 | 13.8 | 13.8 |
| Soybeans | 8.1 | 8.4 | 7.7 | 7.7 |
| Sorghum | 5.9 | 5.9 | 5.5 | 5.5 |
| Wheat | 3.2 | 3.2 | 3.5 | 3.4 |
| Sample size | 405 | 380 | 405 | 380 |

The expected CVs within each year are about the same. The current equal sub-strata sample sizes within strata versus the adjusted optimum unequal sub-strata sample size allocations differ by twenty-five sample units. The 1986 projected cattle CV using rounded unequal sub-strata sample sizes is lower than the projected CV based on equal sub-strata sample sizes. This is due to one sub-stratum in stratum 40 having a very large standard deviation so the sample size is doubled in that sub-strata. Most of the 1987 and 1986 cattle estimates come from stratum 40. Table 6 lists sample allocations used at the sub-strata level for both years. Variation in rounded sub-strata sample sizes within strata was relatively small.

Table 6. Kansas - Strata versus sub-strata allocation analysis
1987 and 1986 sub-strata sample sizes.

| | | Optimal Allocation | | Rounded Allocation | |
|---|---|---|---|---|---|
| Stratum | Sub-stratum | 1987 | 1986 | 1987 | 1986 |
| 11 | 1 | 9 | 14 | 10 | 15 |
| 11 | 2 | 6 | 6 | 5 | 5 |
| 11 | 3 | 6 | 6 | 5 | 5 |
| 11 | 4 | 8 | 9 | 10 | 10 |
| 11 | 5 | 11 | 8 | 10 | 10 |
| 11 | 6 | 11 | 10 | 10 | 10 |
| 11 | 7 | 11 | 10 | 10 | 10 |
| 11 | 8 | 12 | 8 | 10 | 10 |
| 11 | 9 | 6 | 10 | 5 | 10 |
| 11 | 10 | 7 | 5 | 5 | 5 |
| 11 | 11 | 9 | 8 | 10 | 10 |
| 11 | 12 | 15 | 14 | 15 | 15 |
| 11 | 13 | 9 | 11 | 10 | 10 |
| 11 | 14 | 8 | 10 | 10 | 10 |
| 11 | 15 | 8 | 5 | 10 | 5 |
| 11 | 16 | 8 | 7 | 10 | 5 |
| 11 | 17 | 7 | 8 | 5 | 10 |
| 12 | 1 | 8 | 11 | 10 | 10 |
| 12 | 2 | 9 | 10 | 10 | 10 |
| 12 | 3 | 7 | 9 | 5 | 10 |
| 12 | 4 | 10 | 8 | 10 | 10 |
| 12 | 5 | 12 | 13 | 10 | 15 |
| 12 | 6 | 10 | 8 | 10 | 10 |
| 12 | 7 | 10 | 12 | 10 | 10 |
| 12 | 8 | 10 | 10 | 10 | 10 |
| 12 | 9 | 9 | 11 | 10 | 10 |
| 12 | 10 | 8 | 7 | 10 | 5 |
| 12 | 11 | 10 | 6 | 10 | 5 |
| 12 | 12 | 9 | 6 | 10 | 5 |
| 20 | 1 | 8 | 12 | 10 | 10 |
| 20 | 2 | 7 | 10 | 5 | 10 |
| 20 | 3 | 14 | 9 | 15 | 10 |
| 20 | 4 | 9 | 8 | 10 | 10 |
| 20 | 5 | 10 | 11 | 10 | 10 |
| 20 | 6 | 5 | 9 | 5 | 10 |
| 20 | 7 | 14 | 9 | 15 | 10 |
| 20 | 8 | 11 | 8 | 10 | 10 |
| 20 | 9 | 7 | 6 | 5 | 5 |
| 20 | 10 | 9 | 6 | 10 | 5 |
| 40 | 1 | 5 | 10 | 5 | 10 |
| 40 | 2 | 8 | 9 | 10 | 10 |
| 40 | 3 | 7 | 3 | 5 | 5 |
| | | 377 | 368 | 380 | 380 |

## CONCLUSIONS

Strata-level allocations may result in a slightly inflated overall sample size. The slightly inflated sample sizes can be reduced by rounding the strata-level optimum sample sizes down to a multiple of five. The choice of unequal sub-strata sample sizes within strata did not result in any substantial benefit. Overall, 6 percent fewer sample segments produced comparable expected CVs using unequal allocation. Due to the multivariate nature of this study, with a wide range in estimates and CVs, similar results are expected in other major agricultural states. The current practice of performing new frame allocations using stratum level input standard deviations and partitioning sample units equally within each strata is acceptable.

## REFERENCES

Bethel, J. (1986), *An Optimum Allocation Algorithm for Multivariate Surveys*, Staff Report SF&SRB-89, U.S. Dept. of Agriculture, Washington, D.C.

Mergerson, J.W., Clark, M., and Fenley, B. (1986), *Optimum Allocation for Multivariate Surveys: An Improved Implementation*, Technical Note, NASS-AFS-86-01, U.S. Dept. of Agriculture, Washington, D.C.

Mergerson, J. (1988), *A Guide to Computing Multivariate Optimal Allocations*, Internal Documentation, U.S. Dept. of Agriculture, Washington, D.C.