United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Survey Sampling
Branch

NASS Staff Report
Number SSB-90-02

May 1990

COMPUTING DESIGN EFFICIENCY OF AREA SAMPLING FRAMES

James W. Mergerson

Computing Design Efficiency of Area Sampling Frames,

James W. Mergerson, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, May 1990.
Staff Report Number SSB9002.

ABSTRACT

Efficiency measures are computed relative to simple random sampling. Methodology for computing efficiency measures and some computed measures are shown.

KEYWORDS: *Stratified sampling, Design effects, Relative efficiency*

Table of Contents

ACKNOWLEDGMENT: *Ron Fecso* performed a thorough review of this report and provided helpful comments.

SUMMARY

Statistical measures of the efficiency of an area sampling frame are computed using 1988 Missouri June Survey data. The current sampling frame design is compared to simple random sampling at both State and strata levels. This is one aspect of evaluating whether a sampling frame design is efficient. Except for hay, the current design is more efficient than simple random sampling by factors ranging from 3.40 for soybeans down to 1.87 for wheat.

Computing Design Efficiency of Area Sampling Frames

James W. Mergerson

INTRODUCTION

The primary purpose of this analysis is to present measures of the efficiency of a substratified sampling design relative to both a collapsed substrata design and simple random sampling. 1988 Missouri June Agricultural Survey (JAS) data is used as an example.

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 and agricultural similarity. Samples are generally selected with equal selection probabilities within the blocks of the partition.

Each year the Area Frame section of NASS constructs new area sampling frames for, normally, two states. After the frame is used in a survey, an evaluation of the new frame is performed using a set of computer programs referred to as the Area Frame Analysis Package (AFAP). As part of this evaluation it is desirable to compare the performance of the new frame to the performance of the old frame [4]. Also of interest is a measure of the gain in precision by using a geographically substratified sampling design versus an unstratified design with simple random sampling over the entire State. This comparison is called the design effect.

Cochran and Kalton describe the **design effect** as the ratio of the variance of the estimator based on the complex design to the variance of the estimator based on a less complex design. A similar measure is the **relative efficiency.** The relative efficiency is the reciprocal of the design effect.

STATISTICAL METHODOLOGY

Given results from a substratified random sample, an unbiased estimator of v_r, the variance of the mean of a simple random sample from the same population, is given by Cochran. An alternative formula for computing v_r is presented below:

$$v_r = \frac{(N-n)}{(n(N-1))}[\bar{x}_{st} - \bar{y}_{st}^2 + v(\bar{y}_{st})] \tag{1}$$

where

N - overall total number of units

n - overall number of sample units

$\bar{x}_{st} = \sum_{h \in I} N_h \bar{x}_h / N$

I - index set $\{1,2,3, \cdots ,L\}$

L - total number of substrata

N_h - total number of units in substrata h

$\bar{x}_h = \sum_{j=1}^{n_h} x_{hj} / n_h$

n_h - number of sample units in substrata h

$x_{hj} = y_{hj}^2$

y_{hj} - reported amount from sample unit j in substrata h

$\bar{y}_{st} = \sum_{h \in I} N_h \bar{y}_h / N$

$\bar{y}_h = \sum_{j=1}^{n_h} y_{hj} / n_h$

$$v(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h \in I} N_h (N_h - n_h) \frac{s_h^2}{n_h} \tag{2}$$

$s_h^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2$

The estimated overall design effect ($deff_o$) is computed as:

$$deff_o = \frac{v(\bar{y}_{st})}{v_r} \quad .$$

A measure of the gain in precision of substratification within each stratum can also be computed. Equations (1) and (2) are modified to perform computations at the stratum-level. For each stratum, stratum-level quantities are substituted for population quantities and the index set becomes the set of substrata within a stratum:

$$v_{r_s} = \frac{(N_s - n_s)}{(n_s(N_s - 1))}[\bar{x}_{st_s} - \bar{y}_{st_s}^2 + v(\bar{y}_{st_s})] \qquad (3)$$

$$v(\bar{y}_{st_s}) = \frac{1}{N_s^2}\sum_{h \in I_s} N_h(N_h - n_h)\frac{s_h^2}{n_h} \qquad (4)$$

where

N_s - total number of units in stratum s

n_s - number of units in sample in stratum s

I_s - index set $\{1,2,3, \cdots ,l_s\}$

l_s - number of substrata in stratum s

$$\bar{x}_{st_s} = \frac{\sum_{h \in I_s} N_h \bar{x}_h}{N}$$

$$\bar{y}_{st_s} = \frac{\sum_{h \in I_s} N_h \bar{y}_h}{N}$$

The estimated stratum-level design effects ($deff_s$) are computed as follows:

$$deff_s = \frac{v(\bar{y}_{st_s})}{v_{r_s}} \quad .$$

A COMPUTATIONAL EXAMPLE

The current Missouri area frame was developed and first used in the 1988 June Agricultural Survey (JAS). The new frame is stratified according to land-use. Each land-use stratum is further geographically substratified. Geographic substratification is described by Geuder. Strata definitions and other design information are given in table 1. Since strata definitions are primarily based on intensity of cultivation, planted acreages of the following items were selected to be included in the analysis: corn, soybeans, sorghum, winter wheat and all hay. Estimates and coefficients of variation (CVs) for the five analysis variables are presented in table 2. Estimates and variances as a percent of total are presented in table 3 at the stratum levels.

Table 1. Missouri Area Sampling Frame Design Information

Stratum	Number of Units in Population	Number of Units in Sample	Number of Substrata	Stratum Definition
11	13425	130	13	75% or more cultivated
12	9781	90	9	50-74% cultivated
20	14383	90	9	15-49% cultivated
31	3683	10	2	Agri-urban
32	1413	5	1	Residential commercial
40	15272	60	6	Range and pasture
50	33	2	1	Non-agricultural
TOTAL		387		

Table 2. Missouri 1988 June Survey Indications

Crop	Estimate (1,000 ac.)	Coefficient of Variation (%)
Corn	2,200	5.3
Soybeans	4,300	4.0
Sorghum	400	12.4
Winter wheat	1,800	6.4
All hay	3,400	7.9

Table 3. Estimates and Variances as Percent of Total, 1988 Missouri JAS

Stratum	Corn		Soybeans		Sorghum		Wheat		Hay	
	%E	%V	%E	%V	%E	%V	%E	%V	%E	%V
11	54	44	63	52	73	80	53	61	8	2
12	23	27	23	23	15	10	21	12	16	5
20	21	27	13	24	9	6	23	22	31	13
31	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0
40	2	2	1	1	3	4	3	5	45	80
50	0	0	0	0	0	0	0	0	0	0

%E - Stratum level % of State level estimate   %V - Stratum level % of State level variance

State and stratum level design effects and relative efficiences are presented in tables 4 and 5. There are two sets of State level comparisons. The current design, (land-use strata with geographic substrata and random sampling within substrata - **LGS**), is compared to a simple random sampling design (**SRS**). An alternative design, (land-use strata with random sampling within land-use strata - **LS**), is also compared to a simple random sampling design. Table 5 shows the current design compared to a simple random sampling design within strata 11, 12, 20 and 40. The notation LGS_i in table 5 indicates the current stratum i design, where $i \in \{11,12,20,40\}$.

Table 4. State-Level Design Effects - Missouri Area Frame (Relative Efficiencies)

State-Level Comparisons	Corn	Soybeans	Sorghum	Wheat	Hay
LGS vs. SRS	0.43 (2.34)	0.29 (3.40)	0.51 (1.94)	0.53 (1.87)	1.11 (0.90)
LS vs. SRS	0.55 (1.82)	0.36 (2.77)	0.63 (1.59)	0.60 (1.67)	1.20 (0.83)

Table 5. Stratum-Level Design Effects - Missouri Area Frame (Relative Efficiencies)

Stratum-Level Comparisons	Corn	Soybeans	Sorghum	Wheat	Hay
LGS_{11} vs. SRS	0.66 (1.52)	0.77 (1.30)	0.79 (1.27)	0.87 (1.15)	0.94 (1.07)
LGS_{12} vs. SRS	0.95 (1.05)	0.84 (1.20)	0.96 (1.04)	0.87 (1.15)	0.84 (1.19)
LGS_{20} vs. SRS	0.86 (1.16)	0.91 (1.09)	0.98 (1.02)	0.93 (1.07)	0.82 (1.23)
LGS_{40} vs. SRS	0.97 (1.03)	1.00 (1.00)	1.02 (1.00)	0.92 (1.02)	0.99 (1.01)

Results indicated the Missouri area frame sampling design at both State and stratum levels is better than the alternative designs considered. These results are to be expected from stratified sampling, the point here is to quantify the effect for future comparisons. Overall, the design is most efficient for soybeans and corn and is least efficient for all hay. One interpretation of the design effect is in terms of increased sample size requirements for estimates from a simple random sampling design to be as precise as those from the current design. For example, to estimate corn with the same precision using a simple random sampling design would require 2.3 times as many sample units.

At the State level the frame is much less efficient in estimating hay, than for the other four items. Any effort to improve efficiency in estimating hay in Missouri should be focused on stratum 40. Stratum 40 accounts for 45% of the hay estimate and 80% of the variance of the estimate (table 3). Reallocation analysis give some support for an increased stratum 40 allocation (table 6).

Table 6. Missouri Area Frame Allocation Comparison

Stratum	Number of Substrata	1988 JAS Number of Sample Units	1988 JAS Based Optimal Allocation
11	13	130	134
12	9	90	78
20	9	90	90
31	2	10	2
32	1	5	2
40	6	60	85
50	1	2	2
TOTALS		387	393

CONCLUSIONS

The efficiency of stratum 40 is a concern. Missouri has the largest all hay acreage of any state. At the state level, the Missouri harvested hay acreage is second to soybean acreage.

Possible stratum 40 modifications should be explored.

One possibility is to divide stratum 40 into two new strata. One stratum would consist of heavily forested land. The other would consist of hay and pasture land. The forested land should have two square mile segments and the hay and pasture land should have one square mile segments. If these modifications prove not to be feasible, the stratum 40 allocation should be increased.

The summary system should be modified to provide design effect computations.

Design effect computations from the summary system examined overtime could indicate the degeneration of a frame and thus determine which states should be given priority for a new area sampling frame. The design effect computations could also be used to detect ineffective stratification.

When evaluating new area sampling frames, the design effect should be computed relative to both the old and new frame. The evaluation should consider all crop and livestock items, economic variables, land in farm and number of farms as well as cost/benefit analysis of the new frame compared to the prior frame.

REFERENCES

[1] Cochran, W. (1977) *Sampling Techniques*. New York: John Wiley and Sons.

[2] Geuder, J. (1984) *Paper Stratification In SRS Area Sampling Frames;* SF&SRB Staff Report No. 79, U.S. Department of Agriculture, National Agricultural Statistics Service, Washington, D.C.

[3] Kalton, G. (1988) *Introduction to Survey Sampling*. Beverly Hills: SAGE.

[4] Tauchen, R. (1988) *Evaluation of Missouri's New Area Frame and Sample;* Internal Memo, U.S. Department of Agriculture, National Agricultural Statistics Service, Area Frame Section, Fairfax, Virginia.