#12

THE USE OF INTERPENETRATING SAMPLING IN AREA FRAMES

Sample Survey Research Branch
Research Division
Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C.

May 1974

THE USE OF INTERPENETRATING SAMPLING IN AREA FRAMES

By

William L. Pratt

Sample Survey Research Branch
Research Division
Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C.

May 1974

CONTENTS

ACKNOWLEDGEMENTS

I would like to thank Norman Beller, Harold Huddleston, Galen Hart, Benjamin Klugh and Margaret Weidenhamer for their many helpful comments and suggestions in preparing this report. A special thanks is in order for Mary Gorham for the typing of this report and Kathy Tennyson for the graphics.

SUMMARY AND CONCLUSIONS

Data from the Nebraska 1973 June Enumerative Survey were studied to determine the effectiveness of an interpenetrating sample design. Analysis of Nebraska data indicates that the interpenetrating design holds much promise for area sampling. Variances at the State level were well below those that could be expected by treating a systematic sample as a simple random sample. The design assures good sample dispersion and provides unbiased estimates of the variances. Flexibility is another desirable feature of the interpenetrating design. It offers the possibility of testing the effectiveness of different questionnaire designs and enumerator training techniques. Sample allocation can easily be adjusted to increase precision and achieve true rotation to relieve respondent burden. Also as a by-product of rotation, additional samples become available for other uses.

A comparison of variances for systematic selection within replication with those expected from simple random selection of sample units within geographic sub-strata, normally called paper strata, indicates no particular advantage for either method.

Geographic stratification of the frame, provided by the paper strata, was particularly effective in areas of intensive cultivation. In these strata, the analysis indicates the additional geographic stratification provided by paper strata was generally effective in reducing the variance of the estimate for all items considered in this analysis; namely, for hogs, cattle, corn, wheat and soybeans.

It appears that there should be a minimum of 10 paper strata in land use stratum 11 (high intensity cultivated land). Additional geographic stratification provides further gains for most items; however, it is doubtful if more than 20 paper strata would be beneficial as the gain would probably be offset by the loss in degrees of freedom.

Paper stratification in land use stratum 12 (medium intensity cultivated land) was effective as shown by the fact that means of all items were significantly different. The design actually used consisted of 8 paper strata and additional gains from more geographic stratification could be expected if the sample size had been larger. With an allocated sample size of 48 segments it is doubtful that one would want to increase the number of paper strata.

Paper strata in land use stratum 40 (grazing land) contributed very little to reducing variance except for cattle. This is as one would suspect. The additional geographic stratification provided little if any gain in stratum 50 (low intensity cultivated land).

The F-tests showed no significant differences between replications. This is required in order to minimize fluctuation of estimates caused by rotation.

The use of non-contiguous counties for creating paper stratification did not show appreciable gains over the use of contiguous counties. However, county ranking based on a model containing important agricultural variables in a State would be helpful in ordering contiguous counties for paper stratification. In many instances even when ordering contiguous counties, decisions must be made as to which of two adjacent counties should be included next in the paper stratum. Data from a ranking model would be helpful in making this decision.

RECOMMENDATIONS

The performance of the interpenetrating design in the Nebraska area frame seems to justify the use of similar sampling designs in other States.

The selection of sample units should be done using simple random sampling within paper strata to reduce the possibility of grouping of segments between replications. Random selection within each paper stratum would lead to the creation of substrata within land use strata.

Analysis should be done on a design with randomly selected segments within each paper stratum to check the efficiency of this modification.

Increased efficiency and ease of analysis for a replicated area design indicates that a like methodology should be investigated for sampling list frames.

INTRODUCTION

Common Methods of Area Sample Unit Selection

The area sample used by the Statistical Reporting Service is an important source of agricultural data. The advantage of the area frame for sampling purposes is its completeness. In other words, all the land area in the United States can be accounted for by using maps and aerial photography. A complete frame is highly important since unbiased estimates of population values depend on sampling from a complete frame.

SRS is currently stratifying the area frame on agricultural land use. When stratification is complete a probability sample of area segments will be selected. This report will consider alternative sampling plans that can be used to select a sample.

One sampling scheme available is simple random sampling within land use strata. Simple random sampling is a method of selecting n units out of N units, where $n \leq N$, such that all possible samples of size n, drawn without replacement, have an equal chance of being selected.

Another possible method is systematic sampling within land use strata using a single random start. To select a systematic sample of n units out of N units, where $n \leq N$, a unit is selected at random from the first k units and every k^{th} unit thereafter, where $k=N/n$.

SRS has used systematic sampling for a number of years. The primary advantage of systematic sampling over simple random sampling for area sample selection is that dispersion of the sample over the population is assured, because the sampling units in the frame have been ordered by adjacent counties within each State. Dispersion is particularly important when estimates for a large number of items are needed, many of which may be highly concentrated in small geographic areas. In this case systematic sampling will provide more precise estimates of population parameters than simple random sampling.

On the other hand, the primary disadvantage of systematic sampling is that an unbiased estimate of the variance does not exist when the sample has been drawn using a single random start.1/ If the population sampled is in essentially random order an approximation of the variance may be obtained by assuming simple random sampling. This approximation of the variance will be biased generally upward if the population is not random since the sample elements will be correlated. One systematic sample does not provide an estimate of the covariance.

Another method of sample selection is interpenetrating or replicated sampling. The technique consists of drawing r samples or replications, where $r \geq 2$, of size k from N units in the population using the same selection procedures for each replication. Then $r \cdot k = n$, where n is the total sample size. Sampling within replications may be systematic or random.

1/ Morris H. Hanson, William N. Hurwitz and William G. Madow, Sample Survey Methods (New York: John Wiley & Sons, Inc., 1953), Volume 1, page 505.

The Interpenetrating Design with Systematic Selection

A selection procedure using interpenetrating sampling with systematically selected replications from an area frame is considered for purposes of illustration. Prior to sample selection the number of segments to be selected in each stratum is determined, based on cost, variance and other factors.

Each stratum is split into count units. A count unit is a specific area of land with an assigned number of sampling units. The number of sampling units assigned to a count unit is the quotient of the area in the count unit divided by the expected segment size. The number of sample units is rounded to a whole number for the count unit. Count units in a stratum are grouped by counties. Counties are ordered in a manner to preserve geographic proximity with adjacent counties that appear to be agriculturally similar being placed together.

After the number of segments has been allotted to each land use stratum, the number of replications and paper strata in each land use stratum must be determined. Paper strata may be defined as a group of contiguous count units (or sampling units) thereby creating geographic stratification. A list is compiled of the ordered count units in a land use stratum, the number of sample units each count unit contains and an accumulated total of sampling units in the stratum. The count units in a land use stratum are grouped into paper strata, each containing an equal number of sample units.

The number of paper strata (k_i) is equal to the cluster size of each replicate and the sampling interval is N_i/k_i where N_i is the total number of segments (or sampling units) in the i^{th} stratum.

If n_i = number of segments allotted to the sample in the i^{th} stratum,

r_i = number of replications allotted to the i^{th} stratum,

k_i = number of paper strata allotted to the i^{th} stratum,

Then $n_i = r_i \times k_i$ or $k_i = n_i/r_i$.

If systematic selection within replications is desired for stratum i, then r_i random numbers will be selected in the first paper stratum. Selection of segments in other paper strata will be determined by adding a sampling interval to the random numbers selected in the first paper stratum. This procedure results in only r_i random samples (or total degrees of freedom available for error) rather than n_i corresponding to the total number of segments in the i^{th} stratum. Sampling in other strata is done in a similar manner.

The interpenetrating design offers several advantages over one single systematic sample previously used by the Agency. Replicated systematic sampling permits the computation of unbiased estimates of the sampling errors from the sample data and maintains the ease of the systematic selection technique. Sample dispersion is assured; however, the design gives somewhat less control on where the segments fall than with a single systematic sample. Another feature of the design is the creation of paper strata which provides geographic stratification in addition to land use stratification. The design is more flexible than using a single systematic sample for modifying the survey design and makes reallocation of the sample possible at any time without a complete redraw. Sample rotation may be varied from stratum to stratum and achieved by deleting complete replications. Additional samples will become available to increase sample size of a given survey or to create multiple samples as a by-product of rotation.

Systematic selection within replication assures large variation within replication since a replication is composed of one segment from each paper stratum. This procedure should also assure a small variation between replications.

Minimum variation between replication is desirable for two reasons: 1) rotation of segments is accomplished by deleting complete replications, hence replications must be uniform to avoid major changes in the survey level of estimates from year to year when new replications are rotated into the sample; 2) a single systematic sample or replication which has been located with a random start may be viewed as a random selection of a cluster of sampling units from the population of k cluster units where k = N/I (N = Number of sample units in the population and I = Interval of selection).2/ Cochran states that systematic sampling is more precise than simple random sampling if the variance within the systematic sample is larger than the variance of the population as a whole. If there is little variation in a sample relative to the population, the units in the sample are repeating more or less the same information.3/ Hence, by using interpenetrating samples with systematic selection we would expect a considerable gain in efficiency over a simple random sample since good sample dispersion is assured.

A Study--The Nebraska Case

SRS first used interpenetrating sampling in the Nebraska area frame with sampling units (segments) selected systematically within replication.

Data from the Nebraska 1973 June Enumerative Survey were studied to determine the effectiveness of an interpenetrating sample design. The balance of this report will deal with these data.

The Nebraska land use frame consisted of 7 strata. (See Appendix A for definition of strata). The following table gives a breakdown of the complete Nebraska design.

STRATA		NUMBER OF		
NUMBER	DESCRIPTION	SEGMENTS	REPLICATIONS	PAPER STRATA
11	High intensity cultivated land	180	9	20
12	Medium intensity cultivated land	48	6	8
21	Agri-urban	8	2	4
22	Urban	6	2	3
30	Non-agricultural land	2	2	1
40	Grazing land	40	4	10
50	Low intensity cultivated land	40	4	10

2/ William G. Cochran, Sampling Techniques, (New York: John Wiley & Sons, Inc. 1963), page 208

3/ Ibid., page 210

The Design used vs. Other Common Designs

A portion of this anlaysis makes use of the assumption that the variance computed within land use strata using the variance formula for a simple random sample is equivalent to the variance which would be obtained from a simple random sample within the land use strata.

Tract data at the segment level for the following items were used in this analysis: Cattle and calves, hogs and pigs, corn, wheat and soybeans planted for all purposes. These items are among the most important for Nebraska

Variances were calculated for five items between replicated systematic samples as follows:

$$V_1(\hat{Y}) = \sum_{i=1}^{s} \frac{N_i^2}{r_i} \frac{\sum_{m=1}^{r_i} (\bar{X}_{i.m} - \bar{X}_{i..})^2}{r_i - 1}$$, where

$\hat{Y}$ = estimated State total for an item based on the direct expansion,

X_{ijm} = segment total for the m^{th} replication in the j^{th} paper stratum in the i^{th} stratum, where $i = 1,2,...,s$; $j = 1,2,...,k_i$; $m = 1,2,...,r_i$;

This notation will be used through out.

$$\bar{X}_{i.m} = \frac{\sum_{j=1}^{k_i} X_{ijm}}{k_i}$$ = mean per segment in the m^{th} replication in the i^{th} stratum,

$$\bar{X}_{i..} = \frac{\sum_{j=1}^{k_i} \sum_{m=1}^{r_i} X_{ijm}}{n_i}$$ = mean per segment in the i^{th} stratum,

N_i = number of segments in the population in the i^{th} stratum,

r_i = number of replications in the sample for the i^{th} stratum,

k_i = number of paper strata in the i^{th} stratum.

Variances were also calculated for these items within paper strata or geographic stratification as follows:

$$V_2(\hat{Y}) = \sum_{i=1}^{s} \frac{R_i^2}{r_i} \frac{\sum_{j=1}^{k_i} \sum_{m=1}^{r_i} (X_{ijm} - \bar{X}_{ij.})^2}{r_i - 1}$$

, where

$$\bar{X}_{ij.} = \frac{\sum_{m=1}^{r_i} X_{ijm}}{r_i}$$ = mean per segment in the j^{th} paper stratum in the i^{th} stratum,

$$R_i = \frac{N_i}{k_i}$$ = number of replications in the population in the i^{th} stratum.

Both variances, $V_1(\hat{Y})$ and $V_2(\hat{Y})$ are unbiased.

Variances were also calculated ignoring replications and paper strata. The segments were then treated as a simple random sample from the land use strata for variance calculations. For this purpose:

$$V_3(\hat{Y}) = \sum_{i=1}^{s} \frac{N_i^2}{n_i} \frac{\sum_{j=1}^{k_i} \sum_{m=1}^{r_i} (X_{ijm} - \bar{X}_{i..})^2}{n_i - 1}$$

The method of calculating $V_3(\hat{Y})$ is unbiased only if the initial assumption holds.

Variances by stratum were calculated and summed to obtain variances at the State level for tract expansions. Extreme operator totals for cattle and hogs were included. One extreme operator on the list had not been removed from the tract during the summary of the survey in June and this data was deleted for our analysis.

Table 1: Direct Expansions, Standard Errors and Relative Standard Errors for Three Methods of Sampling

State Variance Within Land Use Strata

Item	Estimate	Simple Random		Geographic Stratification 1/		Replicated Systematic	
		Standard Error	Relative S.E.	Standard Error	Relative S.E.	Standard Error	Relative S.E.
Cattle	7,929,315	424,966	5.36	398,117	5.02	381,201	4.81
Hogs	3,516,645	347,152	9.87	324,193	9.22	295,399	8.40
Corn	5,924,378	346,511	5.85	275,322	4.65	320,915	5.11
Wheat	2,955,547	258,145	8.73	202,844	6.86	178,216	6.03
Soybeans	1,098,983	138,063	12.56	107,388	9.77	112,237	10.21

1/ Paper strata which were constructed to represent small geographic areas.

Standard errors for both geographic stratification and replicated systematic sampling are well below the standard errors that would have been obtained assuming simple random sampling within land use strata. Standard errors for corn and soybeans, assuming geographic stratification were considerably lower than those for the replicated systematic design. Sampling errors for replicated systematic sampling were somewhat lower for cattle, hogs and wheat. The Nebraska design was actually a form of cluster sampling with each systematically selected replication forming a cluster.

These results indicate that replicated systematic sampling may be more efficient for some commodities than geographic stratification (i.e. paper strata) and geographic stratification is probably more efficient for others. There seems to be little evidence to indicate that either method is superior to the other. Hence, factors other than sampling efficiency may be considered in determining whether to use systematic selection or simple random sampling within replication for future designs.

One disadvantage of systematic selection across paper strata is the possibility of segments in two or more replications falling close together. If the random starts in the first paper strata fall close together then all segments in these replications will be close. Although theoretically there is nothing wrong with this, it may be desirable to minimize this possibility by selecting independent samples in each paper stratum. This could be accomplished by using a random selection of sampling units within paper strata and assigning the first segment selected in each paper stratum to replication one, the second segment selected to replication two, etc. for rotation purposes.

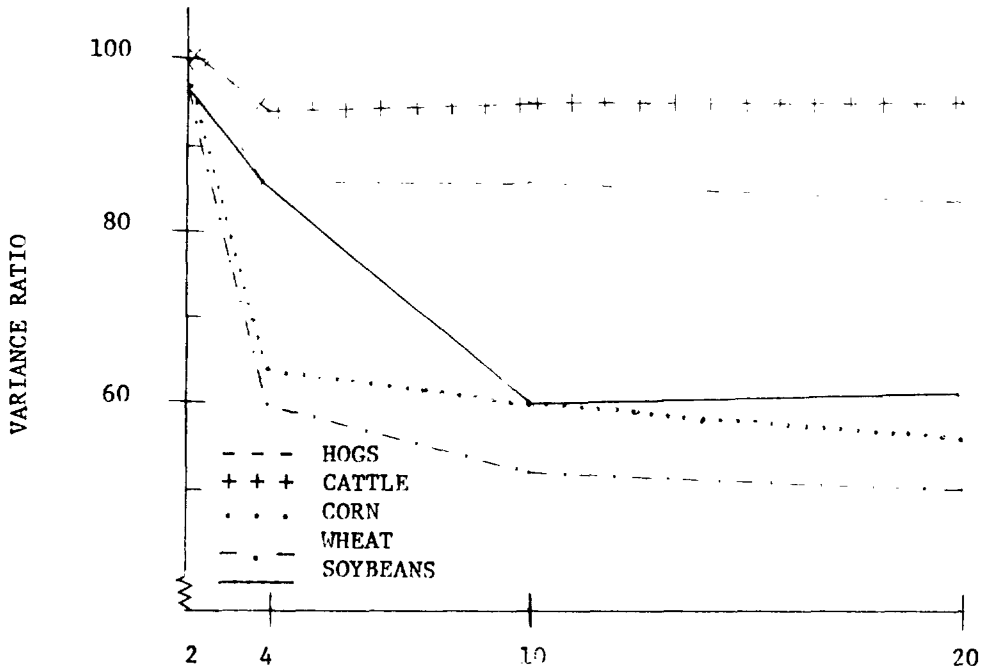
The interpenetrating design was considerably more efficient than simple random sampling would have been within land use strata. The variance formula normally used by SRS for a single systematic sample is the formula for simple random sampling. It is therefore evident that the additional stratification provided by paper strata is quite effective in reducing sampling error as compared with the usual procedure for calculating variance with a single systematic sample.

Analysis to Optimize Numbers of Replications and/or Paper Strata

The second phase of the analysis was to compare the variances of the design actually selected with various other combinations of replications and paper strata that could be created from the Nebraska data. This procedure provides insight into the optimum combination of replications and paper strata. In order to create designs with different combinations of replications and paper strata, adjacent paper strata were collapsed to create larger paper strata. Variances were then computed for various levels of geographic stratification within land use stratum and compared with the variances that might be expected using simple random sampling within land use strata. Segments in strata 11, 12, 40 and 50 were used. Strata 21, 22 and 30 did not contain enough segments to provide meaningful results. An alternative analysis might have been made by assuming the intracluster correlation coefficient would remain constant with a varying number of paper strata and then recomputing the sampling error based on varying the number of paper strata and replicates.

The 180 segments in stratum 11 were combined and regrouped into designs with 2, 4 and 10 paper strata by collapsing the original paper strata. Variances caculated by paper strata showed some improvement over expected variances from simple random sampling for all combinations considered. The largest gains were noted for wheat, corn and soybeans: although, hogs and cattle both showed some gains (See Chart 1).

Chart 1: Ratio V_2/V_3 vs. Number of Paper Strata, Selected Items, Stratum 11, Nebraska--June 1973 4/



NUMBER OF PAPER STRATA
n = 180

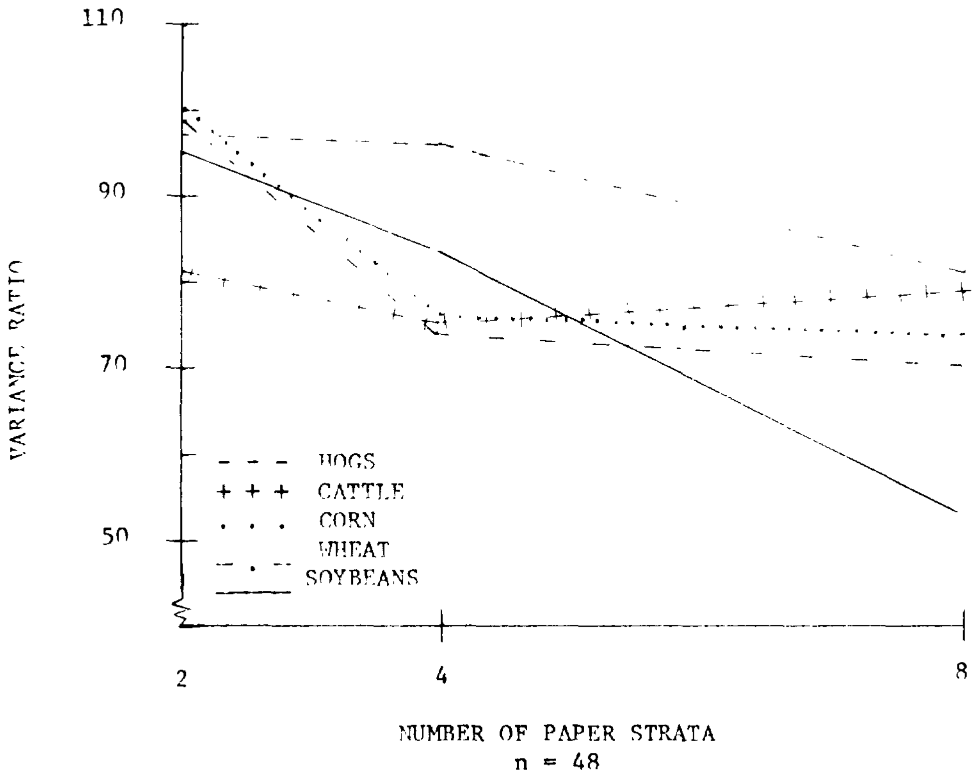
4/ V_2 is the variance for geographic stratification and V_3 is the variance for simple random sampling computed using the formulas on page 5 where i is a fixed value equal to s.

The variances for all commodities were largest when only two paper strata were used. All items showed a marked decrease in variances when the number of paper strata was increased from two to four. The three crop items studied continued to show some decline as the number of paper strata were increased from 4 to 10, while the variance for cattle showed a slight increase. Increasing the number of paper strata from 10 to 20 had very little effect on the variance of cattle. The variance of hogs, corn and wheat continued to decrease, but the variance of soybeans went up slightly.

The greatest reduction in variance for all items occurred as paper strata were increased from two to ten. Reductions in variance for crops were significant, but reductions for livestock were modest. Little variance reduction was obtained by increasing paper strata from 10 to 20.

For stratum 12, variances were calculated for designs of 2, 4 and 8 paper strata (See Chart 2).

Chart 2: Ratio V_2/V_3 vs. Number of Paper Strata, Selected Items, Stratum 12, Nebraska--June 1973

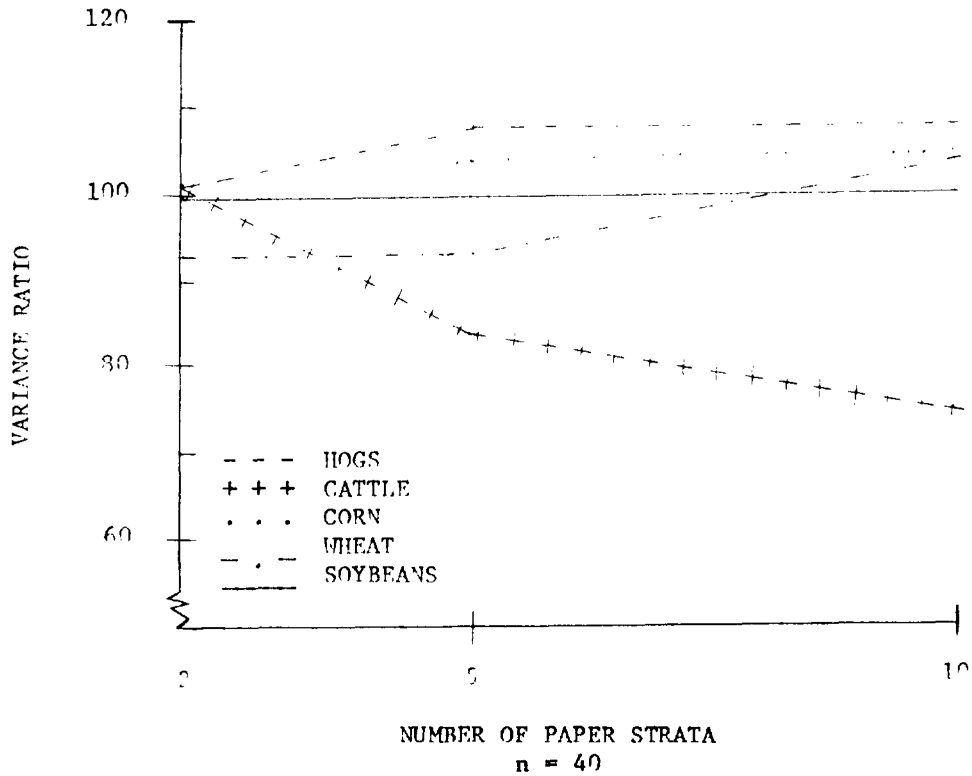


Again, the variances computed were below that of simple random sampling for all items. All five commodities showed a decrease in variance as the number of paper strata were increased from 2 to 4 with the largest decrease for wheat, corn and soybeans. Hogs showed only slight gains.

The increase in the paper strata from 4 to 8 resulted in a modest decrease in variances for corn and wheat while soybeans and hogs showed considerable reduction, reflecting the trend seen in stratum 11.

Combinations of two, five and ten paper strata were considered for land use stratum 40 (See Chart 3).

Chart 3: Ratio V_2/V_3 vs. Number of Paper Strata, Selected Items, Stratum 40, Nebraska--June 1973

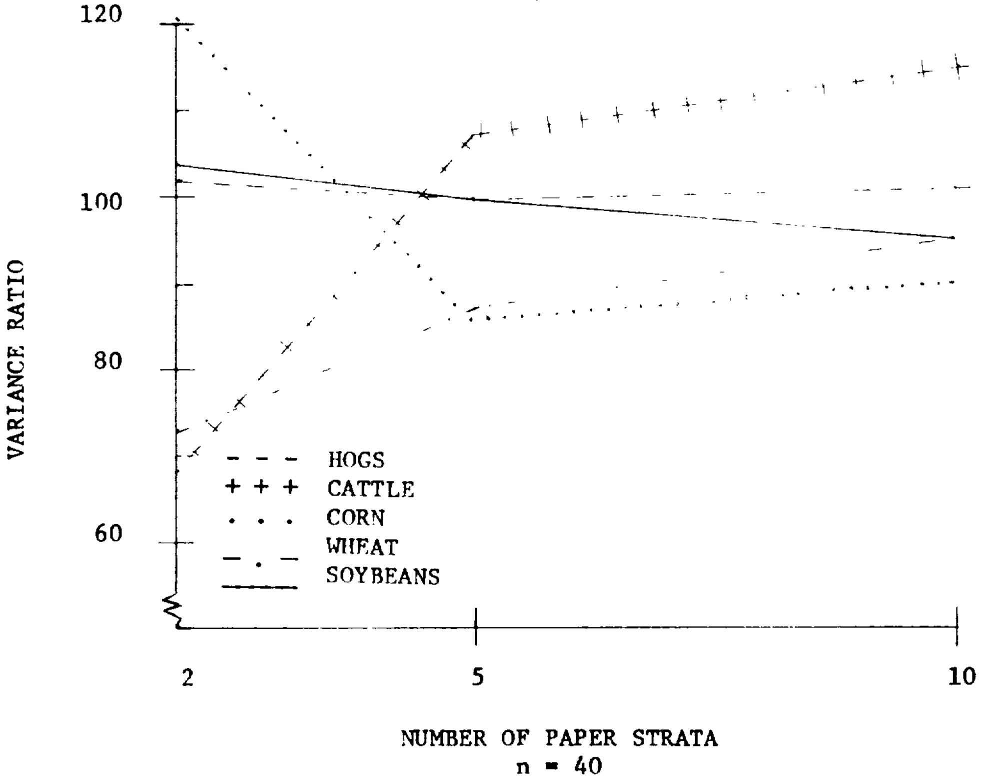


The variance for soybeans was equal to that of simple random sampling for all combinations since all soybeans found in stratum 40 occurred in one paper strata. For two paper strata, wheat showed a slight gain in efficiency over simple random sampling while corn, cattle and hogs were slightly above simple random sampling variances. As the number of paper strata were increased to 5, the variance of wheat remained steadv while cattle showed some improvement. The variance for hogs and corn increased slightly.

When paper strata were increased from 5 to 10, cattle variance continued to show further reduction. However, variances for all other items remained steadv or increased. Stratum 40 is a range stratum and crop items as well as hogs would seldom be found. The benefit of additional stratification within land use stratum 40 is questionable, except for cattle.

Combinations of 2, 5 and 10 paper strata were also considered in stratum 50 (See Chart 4).

Chart 4: Ratio V2/V3 vs. Number of Paper Strata, Selected Items, Stratum 50, Nebraska--June 1973



Some reduction in variances over simple random sampling were noted for cattle, hogs and wheat when two paper strata were used; while the variance of corn and soybeans were above simple random sampling. When number of paper strata were increased from 2 to 5 the variances for cattle, hogs and wheat showed some increase. The variance of corn dropped sharply and soybeans decreased slightly. By increasing paper strata from 5 to 10 all items except soybeans showed an increase in variance. The variances of hogs and cattle for 10 paper strata were above that expected for simple random sampling while the variances for the three crop items were a little below simple random sampling.

Varainces were also calculated between replications for the combinations of replications and paper strata considered. Replications within the collapsed paper strata were renumbered consecutively, starting with the replications in the first paper stratum of those combined and continuing into the replications of the adjacent paper strata. This numbering scheme was somewhat subjective and the new replications were not actually systematic.

Variances were somewhat erratic for the various collapses. Another numbering of the replications that was tried gave completely different results. Therefore, results were not considered to be of sufficient value to be included in this report.

Analysis of Variance Evaluation

A one way analysis of variance by stratum, was calculated for the different combinations of paper strata considered. The analysis of variance procedure was used to test the hypothesis that paper strata means within a land use stratum were equal. Equal paper strata means would indicate that paper strata were not effective in increasing precision over a simple random sample in that loss in degrees of freedom was more than offsetting gains from geographic stratification. See Appendix B for tables.

Analysis of variance indicates that geographic stratification provided by paper strata was generally effective in reducing the variance of the estimates for all items in strata 11 and 12 when more than two paper strata were used. Paper strata means were not significantly different in stratum 40 except for cattle. No significance was found in stratum 50.

Analysis of variance was also used to test the hypothesis of no difference between replication means. No significant differences were found in any stratum for the actual Nebraska design. Results for other combinations of replications and paper strata considered are probably not meaningful due to the subjectivity of reassigning segments to replications as mentioned earlier. See Appendix C for tables.

Multivariate Ranking to Increase the Effectiveness of Paper Strata

The final phase of the Nebraska analysis was to investigate the possibility of improving efficiency of the interpenetrating design by increasing homogeniety of sample units in paper strata. This idea has intuitive appeal for two reasons: 1) reduced variance through more effective substratification, and 2) more uniformity between replications.

An attempt was made to assign an index to each county which would be a measure of the type of agriculture in that county.

Let Y_i = Index number for the i^{th} county

$b_1, b_2, \ldots, b_n$ = Weights for the i^{th} county

$X_1, X_2, \ldots, X_n$ = Agricultural variables for the i^{th} county

Then $Y_i = b_1X_1 + b_2X_2 = \ldots b_n X_n$

Factor analysis was used to determine weights of the variables in multivariate models. This procedure performs a rotation on the variables in such a manner as to explain the largest amount of variance in the system. Weights in the generated eigen vector are then used to form a variable: $Y = b_1X_1 + b_2 X_2 + \ldots + b_nX_n$. The X variables in the model are on a per acre basis for a county and the b-values are weights for the factors. The generated Y values were then ranked and used in combining counties, which may be noncontiguous to form paper strata. In the actual Nebraska design, knowledge of the State was used to group contiguous counties into paper strata.

County data on farmland, value of sales, corn planted for grain and silage, wheat planted, sorghum planted, soybeans planted, all cattle and all hogs were obtained from the 1964 and 1969 U.S. Census of Agriculture. Data for each county were put on a per acre basis in order to eliminate the effect of county size.

Spearman rank correlation coefficients were calculated to obtain correlations between the variables for a particular year (See Table 2 & 3). County rankings by sales were highly correlated with rankings for all variables except wheat and sorghum. Correlations of rankings between like items for the two years ranged between .957 and .986, indicating good stability over time. For any ranking procedure to be feasible, rankings should be stable over time since a frame will probably remain in use for ten years or more.

Data from the 180 Nebraska JES segments in land use stratum 11 were grouped into new paper strata, as closely as possible, to conform with each of the four different ranking models tried. Variances were calculated for cattle, hogs, corn, wheat and soybeans. The procedure was not tried in other land use strata because of the limited number of segments available.

Table 2: Spearman Rank Correlation Coefficients for 1964 Census Data on a per acre Basis

	Sales	Cattle	Hogs	Corn	Sorghum	Wheat	Soybeans
Sales	1.000	.866	.841	.926	.519	.230	.855
Cattle		1.000	.869	.894	.336	-.160	.828
Hogs			1.000	.932	.502	.015	.891
Corn				1.000	.477	.063	.922
Sorghum					1.000	.689	.448
Wheat						1.000	.091
Soybeans							1.000

Table 3: Spearman Rank Correlation Coefficients for 1969 Census Data on a per acre Basis

	Sales	Cattle	Hogs	Corn	Sorghum	Wheat	Soybeans
Sales	1.000	.899	.830	.941	.545	.101	.889
Cattle		1.000	.860	.873	.385	-.207	.817
Hogs			1.000	.853	.568	-.041	.881
Corn				1.000	.524	.040	.893
Sorghum					1.000	.599	.580
Wheat						1.000	.045
Soybeans							1.000

The four models were: 1) farm sales, 2) hogs, 3) sales, cattle, hogs, corn, wheat, sorghum and soybeans, 4) cattle, hogs, corn, wheat, sorghum and soybeans. Variances based on the paper strata formed from the county rankings of the four different models were then compared to the variances of the actual Nebraska design (See Table 4).

Table 4: Variances for Stratum 11, by Paper Strata, Based on Grouping of Noncontiguous counties. (000,000)

		Model			
Item	Actual	1	2	3	4
Cattle	66,572.9	68,313.7	66,785.4	63,866.1	62,796.1
Hogs	72,952.3	80,407.1	72,014.4	71,373.0	74,007.9
Corn	48,458.3	61,162.3	68,814.9	58,453.4	54,911.0
Wheat	22,198.2	21,385.7	26,209.4	21,140.0	20,348.1
Soybeans	10,244.0	11,859.2	14,028.9	12,581.3	11,800.0

County groupings from each model gave a reduction in variance for some items while the variances of other items increased. It was felt that perhaps an aggregate variable such as sales could be used instead of a multivariate model since the ranking by sales was highly correlated with the other variables. However, the ranking by sales was generally worse than other models considered. A grouping was made based on hogs, to see if some reduction could be made in the hog variance over the present sample design. A slight reduction over the actual design was achieved; however, variances for other commodities were considerably worse, as might be expected. The stratification based on hogs produced a larger variance for hogs than the model with the seven variables and showed little improvement over the use of contiguous counties in the actual design.

The other two models failed to produce consistent gains over the actual Nebraska design. It is doubtful if noncontiguous counties can be grouped effectively enough to consistently improve upon paper strata formed using contiguous counties. Geographic stratification appears to be hard to improve upon based on these data.

APPENDIX A

Strata Definitions for the Nebraska Area Frame

Stratum	Intensive Agricultural Land
11	High intensity cultivated land. Almost all land is devoted to crops, intermixed pasture or fallow land. At least 80 percent is cultivated.
12	Medium intensity cultivated land. Land cultivated or subject to cultivation is 50 to 79 percent of the total area. This stratum contains some intermixed pasture or grazing land.
	Towns and Cities
21	Agri-urban. Dwellings and business intermixed with agricultural land.
22	Urban. Dense population, business, industry. No agricultural land.
	Non-Agricultural Land
30	Land unsuited to agriculture or land restricted by law or regulation to a nonagricultural use. This stratum includes parks, waste land, monuments, wildlife preserves or military withdrawals.
	Grazing Land
40	Grazing land. Most is privately owned for seasonal, nomigratory use. State owned lands are probably operated on lease basis with adjacent private land. Within this area are small streams with some hay and/or crops along their banks.
	Extensive Agricultural Land
50	Low intensity cultivated land. Total cropland is between 15 and 49 percent. Fields are too scattered to combine into sensible count units. Grazing land is in smaller parcels than in the grazing stratum.

APPENDIX B

Table 5: One Way Analysis of Variance by Paper Strata for Stratum 11

Number of Paper Strata	Item	Mean Squares (000,000) Between	Within	F	Probability of a greater F
20	Cattle	560.9	369.9	1.52	.0859
	Hogs	1,110.4	405.3	2.74**	.0005
	Corn	2,299.3	269.2	8.54**	.0001
	Wheat	1,276.4	123.3	10.35**	.0001
	Soybeans	402.8	56.9	7.08**	.0001
10	Cattle	726.6	373.3	1.95**	.0475
	Hogs	1,739.4	413.5	4.21**	.0002
	Corn	4,113.8	292.6	14.06**	.0001
	Wheat	2,480.8	127.4	19.47**	.0001
	Soybeans	803.3	56.1	14.33**	.0001
4	Cattle	1,849.8	365.2	5.06**	.0026
	Hogs	4,313.5	414.8	10.40**	.0001
	Corn	10,835.6	308.3	35.15**	.0001
	Wheat	6,047.7	146.8	41.19**	.0001
	Soybeans	842.1	80.9	10.41**	.0001
2	Cattle	36.4	392.3	.01	.9204
	Hogs	130.8	482.1	.27	.6096
	Corn	3,439.3	468.1	7.35**	.0074
	Wheat	1,903.6	236.4	8.05**	.0053
	Soybeans	627.7	90.6	6.93**	.0091

** Indicates Significance at $\alpha = .05$

Table 6: One Way Analysis of Variance by Paper Strata for Stratum 12

Number of Paper Strata	Item	Mean Squares (000,000) Between	Within	F	Probability of a greater F
	Cattle	480.2	175.7	2.73**	.0203
	Hogs	523.8	204.7	2.56**	.0279
8	Corn	985.6	298.6	3.30**	.0074
	Wheat	619.7	161.3	3.84**	.0030
	Soybeans	149.2	21.6	6.91**	.0001
	Cattle	1,028.9	166.0	6.20**	.0016
	Hogs	399.6	242.2	1.65	.1904
4	Corn	1,794.3	305.9	5.87**	.0022
	Wheat	1,100.1	170.2	6.46**	.0013
	Soybeans	144.0	33.6	4.29**	.0098
	Cattle	2,167.6	178.8	12.13**	.0014
	Hogs	546.4	245.9	2.22	.1392
2	Corn	354.6	401.9	.88	.6451
	Wheat	388.5	226.1	1.72	.1936
	Soybeans	125.6	38.7	3.24	.0749

** Indicates Significance at $\alpha = .05$

Table 7: One Way Analysis of Variance
by Paper Strata for Stratum 40

Number of Paper Strata	Item	Mean Squares (000,000) Between	Within	F	Probability of a greater F
	Cattle	3,391.9	1,377.8	2.46**	.0307
	Hogs	8.3	11.9	.70	.7080
10	Corn	105.4	130.7	.81	.6145
	Wheat	89.2	107.0	.83	.5923
	Soybeans	.5	.5	1.00	.4618
	Cattle	4,434.0	1,546.4	2.87**	.0368
	Hogs	3.6	12.0	.03	.8742
5	Corn	85.8	129.3	.66	.6245
	Wheat	165.3	95.8	1.73	.1656
	Soybeans	.5	.5	1.00	.4217
	Cattle	1,357.0	1,855.3	.73	.5978
	Hogs	4.8	11.3	.42	.5261
2	Corn	13.2	127.3	.10	.7486
	Wheat	390.0	95.4	4.09**	.0475
	Soybeans	.5	.5	1.00	.3250

** Indicates Significance at α = .05

Table 8: One Way Analysis of Variance
by Paper Strata for Stratum 50

Number of Paper Strata	Item	Mean Squares (000,000) Between	Within	F	Probability of a greater F
10	Cattle	118.1	309.2	.38	.9345
	Hogs	271.9	305.1	.89	.5449
	Corn	279.8	189.7	1.48	.2019
	Wheat	211.0	173.2	1.22	.3202
	Soybeans	6.2	5.1	1.22	.3223
5	Cattle	97.6	284.2	.34	.8475
	Hogs	345.5	292.0	1.18	.3350
	Corn	466.4	181.2	2.57	.0540
	Wheat	388.8	158.2	2.45	.0629
	Soybeans	5.2	5.3	.98	.5677
2	Cattle	209.1	266.6	.78	.6149
	Hogs	819.7	283.7	2.89	.0937
	Corn	45.0	214.8	.21	.6542
	Wheat	244.6	180.2	1.36	.2500
	Soybeans	13.6	5.1	2.67	.1067

** Indicates Significance at α = .05

APPENDIX C

Table 9: One Way Analysis of Variance by Replications for Stratum 11

Number of Replications	Item	Mean Squares (000,000) Between	Within	F	Probability of a greater F
	Cattle	333.2	392.8	.85	.5624
	Hogs	315.2	487.9	.65	.7397
9	Corn	405.8	488.4	.83	.5775
	Wheat	40.5	255.3	.16	.9948
	Soybeans	63.5	95.0	.67	.7204
	Cattle	296.6	399.9	.74	.7570
	Hogs	254.2	503.8	.50	.9481
18	Corn	280.6	506.1	.55	.9204
	Wheat	104.4	260.5	.40	.9836
	Soybeans	63.3	96.8	.65	.8439
	Cattle	374.0	395.4	.95	.5728
	Hogs	449.2	490.2	.92	.6219
45	Corn	292.1	547.5	.53	.9909
	Wheat	107.5	290.8	.37	.9997
	Soybeans	72.4	100.5	.72	.8950
	Cattle	419.9	360.6	1.16	.2362
	Hogs	499.0	461.5	1.08	.3558
90	Corn	648.9	322.3	2.01**	.0007
	Wheat	271.5	220.2	1.23	.1616
	Soybeans	116.4	71.1	1.64**	.0104

** Indicates Significance at $\alpha = .05$

Table 10: One Way Analysis of Variance by Replications for Stratum 12

Number of Replications	Item	Mean Squares (000,000) Between	Within	F	Probability of a greater F
	Cattle	34.4	243.3	.14	.9795
	Hogs	115.2	268.6	.43	.8270
6	Corn	120.0	434.4	.28	.9227
	Wheat	145.8	239.5	.61	.6959
	Soybeans	21.0	42.9	.49	.7837
	Cattle	127.5	249.7	.51	.8833
	Hogs	204.0	267.0	.76	.6731
12	Corn	223.5	455.1	.49	.8964
	Wheat	175.2	246.1	.71	.7205
	Soybeans	35.6	42.1	.84	.6001
	Cattle	165.5	274.3	.60	.8050
	Hogs	260.8	244.1	1.07	.4360
24	Corn	326.0	472.7	.69	.8117
	Wheat	155.2	300.8	.52	.9415
	Soybeans	37.5	43.5	.86	.6373

** Indicates Significance at α = .05

Table 11: One Way Analysis of Variance by Replications for Stratum 40

Number of Replications	Item	Mean Squares (000,000) Between	Within	F	Probability of a greater F
	Cattle	1,550.3	1,866.9	.83	.5115
	Hogs	16.2	10.7	1.52	.2260
4	Corn	67.7	129.6	.52	.6741
	Wheat	159.4	98.2	1.62	.1998
	Soybeans	.5	.5	1.00	.4053
	Cattle	1,039.0	2,018.3	.51	.8173
	Hogs	13.0	10.7	1.21	.3262
8	Corn	98.2	130.7	.75	.6324
	Wheat	125.3	98.0	1.28	.2916
	Soybeans	.5	.5	1.00	.5499
	Cattle	1,613.1	2,060.6	.78	.7016
	Hogs	17.2	5.4	3.26**	.0061
20	Corn	118.8	130.7	.91	.5813
	Wheat	102.1	103.7	.99	.5112
	Soybeans	.5	.5	1.00	.4985

** Indicates Significance at $\alpha = .05$

Table 12: One Way Analysis of Variance by Replications for Stratum 50

Number of Replications	Item	Mean Squares (000,000) Between	Within	F	Probability of a greater F
	Cattle	86.5	278.0	.30	.8205
	Hogs	367.7	291.6	1.26	.3019
4	Corn	254.4	206.8	1.23	.3126
	Wheat	277.4	173.9	1.59	.2066
	Soybeans	2.8	5.5	.51	.6810
	Cattle	183.0	283.0	.65	.7159
	Hogs	303.4	296.1	1.02	.4339
8	Corn	115.5	231.3	.50	.8286
	Wheat	143.3	190.3	.75	.6315
	Soybeans	4.4	5.5	.79	.6016
	Cattle	354.7	180.0	1.97	.0703
	Hogs	274.0	319.7	.86	.6303
20	Corn	163.1	255.5	.64	.8336
	Wheat	233.7	132.7	1.76	.1091
	Soybeans	5.1	5.5	.92	.5678

** Indicates Significance at α = .05

REFERENCES

1. Deming, W. Edwards, Sample Design in Business Research, New York: John Wiley & Sons, 1960.

2. Cochran, William G, Sampling Techniques, New York: John Wiley & Sons, 1963.

3. Hansen, Morris H., Hurwitz, William N. & Madow, William G., Sample Survey Methods and Theory, Volume 1, New York: John Wiley & Sons, 1953.

4. Snedecor, George W. & Cochran, William G., Statistical Methods, Ames: Iowa State University Press, 1967.