DEVIATING FROM THE OPTIONAL ALLOCATION TO REDUCE BURDEN ON LARGE-SIZED UNITS

David M. Dillard and Barry L. Ford, U.S. Department of Agriculture

ABSTRACT: In a stratified sample design, an optimal (or Neyman) allocation distributes the overall sample size among the strata so that the sampling error is minimized. For skewed distributions, the gains from an optimal allocation are substantial, but the sampling rate for units in the upper tail of the distribution is usually extremely high. This paper discusses how optimal allocations are putting a severe burden on large farms in the U.S. To decrease the burden, the U.S. Department of Agriculture (USDA) has begun deviating from optimal allocations. This paper discusses the method of alternative allocation in its effects.

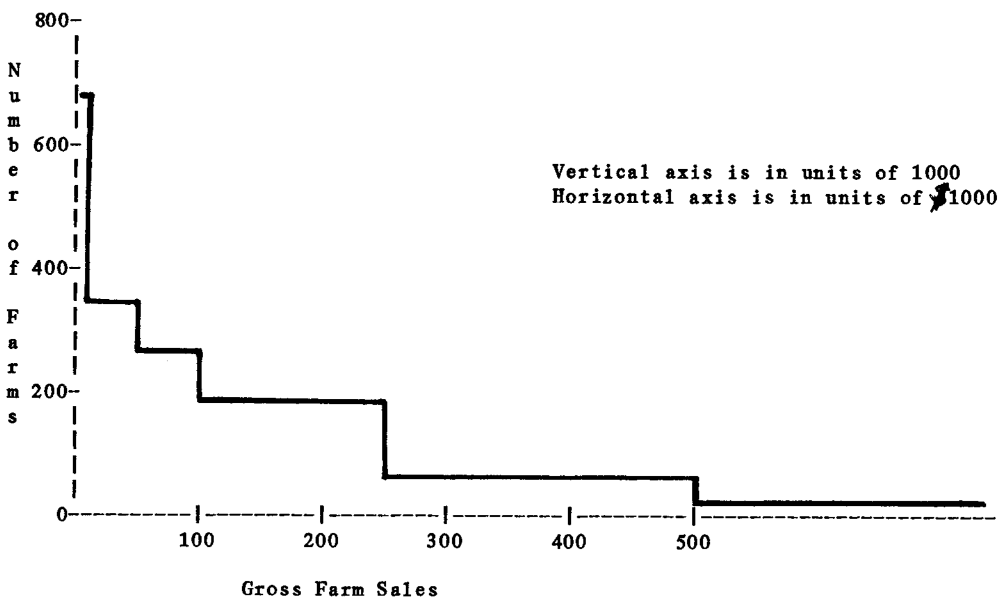
BACKGROUND: An extremely common occurrence in survey design is the population distribution which is skewed to the right (Hansen, Hurwitz, and Madow, pp. 141-147). These distributions result from most of the population clustering in a mass of smaller values while units with substantially larger values trail out to form the upper tail of the distribution. This type of distribution is typical of many size measures -- number of acres in a farm, sales of a retail store, income, etc. -- and is also typical of the variable of interest in this paper -- gross sales of a farm. Figure 1 shows the distribution of U.S. farms by gross sales.

Stratification of skewed populations usually categorize these large-sized units into a stratum (or strata) so that they are separated from the more homogeneous part of the population. In the numerous agricultural surveys run by USDA, stratification usually involves "upper strata" that are composed of the largest farms. Agricultural surveys focus on many farm variables -- gross farm income, number of cattle, harvested acres of cotton, etc. -- but the farms composing the upper strata tend to come from the same general subset of the farm population.

Optimal allocation (Cochran, page 95) is the common method of allocating the total sample size to the individual strata so that the smallest possible standard error results. Optimal allocations can also be formulated to account for different costs that may occur in different strata.

In skewed distributions, the gain from optimal allocation is usually large in comparison to proportional allocation. Table 1 shows the difference in proportional vs. optimal allocation for a survey to collect data on farm costs and income -- the 1984 Farms Costs and Returns Survey. Lists of farmers are stratified using gross farm sales and type of commodities produced. As the

Figure 1. Distribution of the number of farms in the U.S. by gross farm sales, 1984.



reader can see, the first stratum contains the largest farms -- referred to as "extreme operators" by USDA. These farms tend to have over $500,000 in gross farm sales.

The optimal allocation gives a much larger sample size to the stratum of large farms than the proportional allocation because of the great variability among the large farms in relation to the other strata. Thus, the sampling rate is 21% which is also much larger than the rate in the other strata. This rate represents a significant burden for the large farms because the questionnaire for the survey is extremely long -- approximately 25 pages -- and requires very detailed data on a subject about which the farmer is sensitive.

Although the example in Table 1 has a sampling rate of 21% for large farms, many other examples in agriculture would show the upper strata with sampling rates of 50% or 100%. As mentioned earlier, these sampling rates are basically affecting the same group of large farms. Accounting for the whole scheme of agricultural surveys results in a tremendous burden on this subset of farms. The high sampling rates produced by optimal allocations reinforce the notion among survey designers that the large farms need to be heavily sampled because of the large amount of agriculture under the control of these farms. For example, large farms (i.e., farms having over $500,000 in gross farm sales) account for 22% of total farm costs but compose less than 2% of the number of farms. Even when survey designers deviate from optimal allocations, they do so in a

Table 1. Allocations to strata in California for the 1984 Farm Costs and Returns Survey.

Stratum	Population Count	Proportional Allocation	Optimal Allocation	Optimal Sampling Rate*
large farms	1,106	17	234	21.2%
1	9,936	150	23	.2%
2	3,155	48	34	1.1%
3	1,408	21	5	.4%
4	323	5	15	4.6%
5	14,356	216	150	1.0%
6	936	14	10	1.1%
TOTAL	31,220	471	471	1.5%

* Optimal Sampling Rate = (Optimal Allocation / Population Count) x 100%.

Table 2. For three types of allocation, the relationship between the CV of the entire sample and the sample size in the stratum for large farms.

Type of Allocation	CV for Entire Sample	Sample Size for Stratum of Large Farms
Optimum	9.6	234
Operational	10.1	201
Reduced	10.6	138

subjective manner rather than using a precise goal or procedure.

ANALYSIS: The first purpose of the analysis was to explore the relationship between the sample size in the stratum for large farms and the standard error of the entire sample. Rather than referring to the standard error, this paper often refers to the coefficient of variation (CV), which is a percentage that expresses the standard error relative to the estimate itself.

Figure 2 shows the curved line that relates the CV for the entire sample to the sample size in the stratum for large farms. The far right line represents the optimal allocation. The line next to it represents the sample size that was eventually used for the operational survey -- its deviation from the optimal size was somewhat subjective. As the reader can see, the curved line is flat near the line of optimal allocation. Although the operational allocation has moved to a smaller sample size with little effect on the CV, the survey designers could have reduced the sample size even more with little effect on the CV.

By calculating the sample size that results from letting the CV increase 1%, the authors calculated the line called "reduced allocation". (Obviously, the authors could have used any x% change as a criterion.) This method at least gives a formal criterion for deciding how much to deviate from the optimal allocation rather than the subjectivity of the operational allocation.

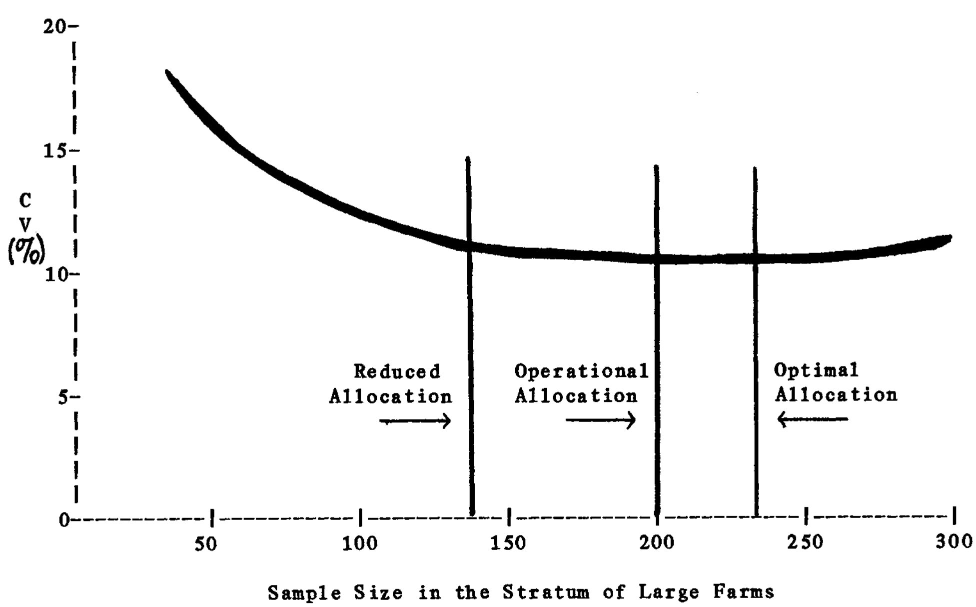
The effects of the three types of allocation on sample size and CV are shown in Table 2. By allowing the 1% increase in CV, the reduced sample size for the stratum of large farms decreased 41%.

Table 3 compares the effect of the alternative allocations across the entire U.S. for the 1984 Farm Costs and Returns Survey. Reduced allocations were never allowed to go below 3 in any one state. Over the entire U.S., there was a 39% decrease in sample size in going from the optimal to the reduced allocation. Although the operational allocations had already decreased 18% from the optimal allocations, there was still much room to decrease the sample size. The effect of a 1% increase in CV for each state is negligible on the national CV.

The biggest decreases in sample size were in states with large optimal allocations. Three states -- California, Kansas and Texas -- accounted for almost half the reduction.

SUMMARY: This paper examines the heavy survey burden put on large-sized units when optimal allocations are repeatedly applied to the same population. In this paper the population of interest is farmers, and the burden is the result of numerous agricultural surveys. By examining the curved line that relates the CV to the sample size in the stratum containing large farms, the authors find that deviating from the optimal allocation is a logical approach to relieving burden. For the farm survey examined in this paper, substantial

Figure 2. Relationship between the coefficient of variation (CV) for the entire sample and the sample size in the stratum of large farms for California.



reductions in sample size result with only small effects on the CV.

The statistical literature recognizes the fact that moderate deviations from the optimal allocation, i.e., "imperfect" allocations, do not have major effects on the CV's (Cochran, pp. 114-115). In other words, the literature recognizes the fact that curves relating CV's and changes from optimal allocations are usually flat, but the literature also assumes that the fpc is negligible. This analysis makes a simple extension of this flatness notion to the upper strata where the fpc's become crucial because optimal sample rates are extremely high. This extension is important only because: 1) the situation arises so frequently, 2) the optimal allocations reinforce the notions of survey designers that the large-sized units need to be heavily sampled, and 3) the situation can have such drastic effects on survey burden. Given the idea that survey designers should deviate from optimal allocations in these situations, the method in this paper is a simple and direct way of deciding how much to deviate.

Some readers will recognize that the problem motivating this analysis -- allocations creating unreasonable burdens --stems from the fact that designers of agricultural surveys are not using a formal measure of respondent burden. If the survey designers could create an index of respondent burden, they could use it either as another cost measure in allocation formulas or as a size measure for doing pps sampling within each stratum (Tortora).

However, the concept of respondent burden across all agricultural surveys is an inexact notion to most survey designers. How would one quantify it? Bradburn has noted that the perception of survey importance by the respondent, the stress upon the respondent by the survey, and the effort required by the respondent to complete the survey are some of the integral components of respondent burden. Attempts have been made to define overall survey burden at USDA, but the results have never been satisfactory. More research should be done on this approach. Until the problems in measuring respondent burden are overcome, the method given in this paper seems a good way of attacking the problem for agricultural surveys.

Table 3. Comparison of Optimal, Operational, and Reduced Allocations in the stratum of large farms for the 1984 Farm Costs and Returns Survey.

State	Population Count	Optimal Allocation	Operational Allocation	Reduced Allocation
Alabama	54	3	6	3
Arizona	163	72	45	53
Arkansas	51	14	12	3
California	1113	234	201	138
Colorado	647	174	111	133
Connecticut	23	9	6	6
Delaware	12	4	3	3
Florida	366	58	48	30
Georgia	1098	33	36	19
Idaho	334	54	45	31
Illinois	25	3	9	3
Indiana	42	3	6	3

Iowa	150	15	18	3
Kansas	238	156	96	76
Kentucky	17	3	6	3
Louisiana	42	9	9	3
Maine	30	3	6	3
Maryland	34	7	9	4
Massachusetts	7	2	3	3
Michigan	139	11	12	3
Minnesota	250	13	33	3
Mississippi	46	19	21	4
Missouri	47	3	9	3
Montana	917	13	18	6
Nebraska	477	82	72	34
Nevada	102	34	24	27
New Hampshire	7	3	3	3
New Jersey	35	3	6	3
New Mexico	1060	92	57	81
New York	212	4	12	3
North Carolina	125	13	21	3
North Dakota	240	5	12	3
Ohio	413	35	48	13
Oklahoma	168	62	30	34
Oregon	337	14	15	8
Pennsylvania	52	12	12	4
Rhode Island	NO LARGE FARMS			
South Carolina	68	5	9	3
South Dakota	423	12	21	3
Tennessee	41	3	6	3
Texas	1757	388	294	239
Utah	124	21	21	10
Vermont	16	5	3	3
Virginia	143	3	6	3
Washington	452	150	87	108
West Virginia	23	10	6	6
Wisconsin	153	14	15	3
Wyoming	415	50	36	39
TOTAL	12,688	1935	1584	1175

REFERENCES

Bradburn, Norman. "Respondent Burden", ASA Proceedings of the Section on Survey Research Methods, pg. 35-40. 1978.

Cochran, William G. SAMPLING TECHNIQUES, John Wiley and Sons. 2nd Edition, 1963.

Hansen, Morris H., Hurwitz, William N., and Madow, William G. SAMPLE SURVEY METHODS AND THEORY, John Wiley and Sons. 1953.

Tortora, Robert D. and Crank, Keith N. "Shifting Respondent Burden Through Unequal Probability Sampling", Estadistica, pg. 71-90. 1980.