Journal of Official Statistics, Vol. 12, No. 1, 1996, pp. 101–104

Comment

Mary H. Mulry[1]

Kadane has made a major contribution to the planning of the 2000 census by framing the approach to sample design. Unlike previous censuses, the plans for the 2000 census call for using sampling at two points in the data collection: (1) the follow-up interviewing of the nonrespondents to the mail questionnaires and (2) the integrated coverage measurement survey. Kadane has been the first to consider the sample allocation in the context of using the census numbers for the apportionment of Congress. Designing the sampling plan for the census is difficult because of the apportionment formula, the nature of census coverage error, and the goal of equity.

Equity, or fairness, is always an issue for allocation of Congressional seats and Federal funds, particularly because a "fixed pie" is being distributed. When the basis for the allocation is data with sampling error, the uncertainty in the numbers makes the issue more complicated. The working assumption is that allocation of the sample should support the fair allocation of resources.

The issue is how to define equity. In essence, the definition of equity is a policy decision. However, there are technical implications for different ways of defining equity. The Constitutional basis of the census is the apportionment of seats in Congress among the states. One approach is to give apportionment the highest priority when assessing the effect of sampling variance in the census numbers. With this line of reasoning, fund allocation and other uses should follow in priority.

Underlying the apportionment algorithms for Congressional representation is the requirement to make specified quantities as equal as possible among the states. The criteria for choosing the elements which are to be made as equal as possible in apportionment may be expressed as loss functions (Balinski and Young 1982; Spencer 1985). In the current apportionment algorithm, Hill's method, the average district sizes are taken as the basic element to be made as equal as possible (Balinski and Young 1982). This is more apparent when the loss function for Hill's algorithm is expressed as on the right side of the equation

$$\sum_i a_i^{-1}(a_i - q_i)^2 = \left(\frac{435}{P_+}\right)^2 \sum_i a_i \left(\frac{P_i}{a_i} - \frac{P_+}{435}\right)^2$$

---

[1] Mary H. Mulry is Mathematical Statistician, Decennial Statistical Studies Division, U.S. Bureau of the Census, Washington, DC 20233, U.S.A. This article reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

where $P_i =$ the population of state i, $i = 1, \ldots 50$

$P_+ = \Sigma_i P_i$

$q_i = 435(P_i/P_+)$, called the quota for state i

$a_i =$ the number of seats allocated to state i.

In contrast, if individuals are considered the basic elements whose shares are to be made as nearly equal as possible, then Webster's method minimizes this type of error. The expression of the loss function for Webster's algorithm on the right side of the equation illustrates

$$\sum_i P_i^{-1}(a_i - q_i)^2 = \sum_i P_i\left(\frac{a_i}{P_i} - \frac{435}{P_+}\right)^2.$$

The Hill method was chosen for political reasons at the time. However, the Supreme Court has ruled it to be fair as recently as 1993 (Ernst 1994).

The sample design to support allocation of resources also may affect the fairness. Some areas or subpopulations may receive more than they would have in the absence of sampling error while other areas or subpopulations may receive less. The choice of the equity definition has a significant effect on the allocation of the sample. Previous work on the effects of sampling error in the allocation of a sample for distributing funds does not apply directly to Congressional apportionment because states receive whole seats, not fractional parts.

Investigating and fully describing the interrelationship between definitions of equity and sampling error is a topic of research. Such a description will facilitate a policy decision on the definition of equity to pursue in allocating the samples in the 2000 census.

When designing a survey that will produce estimates for several entities, statisticians often use relative error as a measure of equity and specify uniform coefficients of variation (CV). Kadane suggests equal CVs as the criteria for sample allocation for the census. However, CVs may not be the appropriate measure to examine for allocating a sample to states when the estimates will be used in the apportionment formula. The apportionment formula is very sensitive to small changes in the number of people in the states that receive the last few seats. The requirement of equal coefficients of variation for each state tends not to produce an apportionment which is robust to the uncertainty in the estimates. A CV of 0.5% for California is a standard error of about 150,000 while a CV of 0.5% for Wyoming is about 2,300. On average, each Congressional district should have about 570,000 people. With a CV of 0.5% the radius of a 95% confidence interval for California (approximately 300,000) would be about half the size of a Congressional district.

A standard error of 150,000 for the estimated size of California is almost sure to cause a lack of robustness in the apportionment process. To illustrate, when a simulation of the apportionment process assumed the populations of the 50 states had a multivariate normal distribution with means, the 1990 Post Enumeration Survey (PES) estimates and CVs of 0.5% with all the covariances set equal to 0, 18 states had at least one error in the number of seats allocated in 100 repetitions. Nine states

Table 1. Number of states with errors in allocation of Congressional seats in 100 repetitions with the 1990 Post Enumeration Survey estimates of population size

	Number of errors		
Allocation of 750,000 housing units	$\geqslant 1$	$\geqslant 10$	$\geqslant 46$
Equal CVs, CV = 0.5%	18	9	1
Proportional to $\sqrt{\text{size}}$ Fix California CV = 0.15%	14	5	1

had more than 10 errors in the 100 repetitions while the most times one state had an error was 47. However, restricting the error in California and using proportional allocation in other states may not be appropriate either. When the *CV* of California was lowered to 0.15% and other states were allocated sample proportional to the squared root of the size of the state, fourteen states had at least one error in 100 repetitions, with the most times one state had an error being 46. In this allocation, all states had a *CV* lower than 1.0% with the exception of five states, of which Wyoming had the highest *CV* at 1.21% (Navarro 1995a). Table 1 demonstrates that there is not much difference in the simulation results when the sample allocation supports equal *CV*s and when it supports proportional allocation.

If the requirement for the *CV* for a direct estimate for each state is 0.5%, then the estimated sample size for the integrated coverage measurement survey is 750,000 housing units (Navarro 1995b). This estimate is based on the estimates of the 1990 PES. Surprisingly, some of the smaller, rural states may require a disproportionately larger sample because the coverage error in these states had larger variability. In these states, the net coverage error is very small, but the gross coverage errors are large. It is not clear whether this phenomenon will be present in the 2000 census data. Possibly this occurrence in 1990 was due to geocoding error. With the improvements in the TIGER database (the automated mapping system) and the development of the Master Address File (MAF), the gross coverage errors in these states may not be as large in 2000. Evaluations of the MAF will, we hope, indicate the extent of the problem prior to 2000, and this information will be taken into account in the sample design.

Direct estimates for states probably will be a necessity because states will not want their population sizes used in the apportionment algorithm influenced by the census in other states as it would be with indirect estimates. Sample allocation within states will be affected by two additional factors: (1) there may be substate areas that also need direct estimates, and (2) coverage error is more variable for some subgroups than for others.

Any sample design must consider that the robustness of the apportionment process appears to depend heavily on the precision of the estimated population of the states receiving the last few seats in the allocation process. An approach that provides some robustness in the apportionment process is to guarantee each state a minimum *CV* and make the estimates for states expected to be the candidates for the last few seats more precise. Possibly, the 1999 postcensal estimates could be used to predict which states need larger samples. The success of this approach would depend on the accuracy of the estimates selected in 1999.

A complementary approach to determining the sample size would be to design the sample to produce a specified accuracy for areas having populations equal to the average size of a Congressional district (Spencer 1995). In Congressional redistrictings that have been challenged in court, the rulings have required that districts have the smallest variation in size practicable, unless there was a compelling state purpose that applied uniformly across the state. There is no minimal variation which is guaranteed to be acceptable. A difference of less than 1% between the largest and the smallest district has been rejected when there did not appear to be a good faith effort to have the smallest difference possible. For state and local redistricting, the courts have allowed a little more variation but limited the deviation in the size of the districts to be at most 10% (Grofman 1990).

Accuracy and equity entwine to present a very interesting challenge for sample allocation and estimation for the 2000 census. Statistical research can describe the relationship between sampling error and the apportionment formula. The choice of the level of error to permit in the design is ultimately a policy decision.

References

Balinski, M.L. and Young, H.P. (1982). Fair Representation. New Haven: Yale University Press.

Ernst, L. (1994). Apportionment Methods for the House of Representatives and the Court Challenges. Management Science, 40, 1207–1227.

Grofman, B. (1990). Voting Rights, Voting Wrongs: The Legacy of Baker v. Carr. New York: Priority Press.

Navarro, A. (1995a). Apportionment Simulation Results. Note to John H. Thompson dated February 1, 1995, Decennial Statistical Studies Division, U.S. Bureau of the Census.

Navarro, A. (1995b). Unpublished manuscript. Decennial Statistical Studies Division, U.S. Bureau of the Census.

Spencer, B.D. (1985). Statistical Aspects of Equitable Apportionment. Journal of the American Statistical Association, 80, 815–822.

Spencer, B.D. (1995). Report on Topics Discussed at the Sampling and Estimation Steering Committee Meetings. Unpublished manuscript, U.S. Bureau of the Census.

Received January 1996