Journal of Official Statistics, Vol. 12, No. 1, 1996, pp. 105–106

Rejoinder

Joseph B. Kadane

I thank both Mary Mulry and Alan Zaslavsky for the careful consideration of the sample design issue I raise. While their perspectives are different from mine and from each other, each is a contribution to a discussion of design for the sampling that will, I hope, be part of the census of 2000.

With respect to Mary Mulry's comment, her simulation reveals two difficulties with current thinking about sampling for apportionment. She takes the 1990 PES results as "true," adds error for each run of the simulation, calculates apportionments, and then asks how many states' apportionments change. As she herself recognizes "that apportionment formula is very sensitive to small changes in the number of people in the states that receive the last few seats" (and also the states that would have received the next few seats). Consequently the results of a simulation such as hers depends on how certain the population numbers are for those states for the 1990 PES numbers. This is unlikely to have much to do with the situation after the census of 2000. Would it not be more enlightening to "center" the population in a random location that changes from one computer run to the next? This is exactly what the Bayesian analysis does, taking the "true" population as unknown but having a distribution.

The second issue has to do with a loss function. To count the number of states with errors, as Mulry does, implicitly uses a loss function that is 1 if a state has an error and is zero elsewhere. This pays no attention to the magnitude of the error, unlike the loss functions discussed in the article. I think it is more reasonable to have lower loss for proportionately small errors and larger loss for bigger ones, as L_1 and L_2 do. Implicitly both loss functions imply that an equal-sized error, say of one seat, is less drastic for a state with many seats than for a state with few seats. One may or may not endorse that consequence of these loss functions but that is their property.

Zaslavsky's comment presents a rather different institutional framework than the one I am using. My assumption is that the law will not change, and hence, after the census of 2000, the Hill algorithm will be applied to the estimates $\hat{\phi}_i$ to apportion Congressional seats among the states. What is new is that the Census Bureau will publish estimated standard errors of $\hat{\phi}_i$ in 2000. The question I am addressing is how the Census Bureau can defend such an apportionment against some state's claim that uncertainty entitles it to another seat.

In this connection, I find puzzling Zaslavsky's statement that "the survey designer is being asked to control the information available to the Bayesian in such a way that the Bayesian will come up with an answer that satisfies (7) without being told what those constraints are." The constraint (7) is the current law, which presumes state populations known with certainty, and applies the Hill algorithm to those populations. My article anticipates stochastic knowledge of state populations, and asks how these two can be reconciled. I find a set of sufficient (but not necessary) conditions for the apportionment currently required by U.S. law and the optimal Bayesian apportionment under a particular loss function to be the same.

Zaslavsky is quite correct that his (5) is not an "if and only if" statement, and thus a wide range of coefficients of variation for Wyoming will produce the same apportionment. The largest states are those most likely to be right around the 435th seat in the allocation order, and for them I think that equal state coefficients of variation is not a foolish idea. It is not too far from what the Census Bureau did in 1990 with the PES, where they designed the sample to achieve equal area coefficients of variation, as discussed by Hogan (1992).

Must this result in equal state sample sizes, as Zaslavsky fears? I think not, because I think that variances of estimated coverage rates are likely to grow with ϕ_i, say proportionally to ϕ_i^a for some $a > 0$. Then the larger states would get larger sample sizes, as seems reasonable. Thus I suspect that departures from Zaslavsky's HVA will be systematic, and in a reasonable direction. This is an empirical issue, which perhaps a reanalysis of the 1990 data might help elucidate.

A more serious threat to a soundly designed census sampling plan lies in the idea, alluded to by Mulry, that Congress will insist on methods that depend only on sample results for each state, thus barring methods that borrow strength across states. This leads to a very inefficient design, or, put another way, requires a larger sample size, and hence a larger budget, for the same accuracy. Unfortunately, the U.S. Congress is likely to insist on a my-state-only estimation plan, and then not to appropriate the money needed to increase the sample sizes accordingly.

The design of sampling in conjunction with the U.S. decennial Census of 2000 is an important issue where public policy and technical statistical considerations both matter. I thank both of my discussants for joining me in giving their perspectives on it.

Received March 1996