POLIDATA Political Data Analysis

REMARKS OF
CLARK H. BENSEN

CENSUS 2000 ADVISORY MEETING
U.S. BUREAU OF THE CENSUS

26 SEPTEMBER 1996
WASHINGTON, DC

December 31, 2000. The President will issue the official count of the nation and promulgate the number of seats each state will be assigned in the U.S. House for the next ten years. During the years 2001-2002, an important process will take place in each of the 50 states, a traditional event called redistricting. This activity occurs in 99 state legislative chambers and results in over 7,000 legislative districts and all but a handful of the districts in the U.S. Congress. It is important to remember that it is political representation, specifically the apportionment of seats in the U.S. House, and therefore the electoral college, which raises concern over the census to a constitutional level.

During this redistricting process, thousands of persons are involved and millions of dollars are expended. Preparation for this complex undertaking is already underway in some states. The process is fraught with an abundance of political interests to frustrate any legislator. Moreover, the redistricting process is highly regulated by court decisions which affect the population mix of any district as well as the total population of each district. Suffice to say that when it comes to the population in any district, there needs to be an *absolutely equal* number of persons in each district, based upon the census counts.

For reapportionment, the mathematical formula used to decide which state gets the last few seats in the U.S. House, is very sensitive to small shifts in population. Everything is relative in the method of equal proportions. Whether California gets another seat at the expense of Wisconsin, for example, could be determined by a few thousand persons.

The level at which the entire redistricting process takes place varies from state to state. However, whereas some districts may be created by a combination of whole counties, or whole townships or cities, it is more often than not that the lowest level

used is the census block. This is certainly the case in large urban areas. It has been increasingly the case in otherwise rural districts which include a portion of an urban area.

It is this atomic level of the census block which concerns me the most. As one who has been involved with redistricting since the 1970s, and as one responsible for the creation of the districts, or the datasets used to analyze them, the use of sampling and estimation concern me greatly. Moreover, the results of the 1995 Test Census give rise to even greater concern about the feasibility of sampling and estimation and the reliability of the numbers which will be delivered by the Bureau following Census 2000.

As my handout indicates, the use of sampling and estimation in the census results in a measurable error. The error reflects the fact that the sample which was used may not be representative of the entire universe. As such it quantifies the degree of difference between the numbers generated by the sampling and estimation technique and numbers which would have been generated by a complete count. These standard error figures can be used to create a confidence level for the user. In essence, a confidence level of 95% means that in 95 out of 100 different samples, the sample was representative of the full universe which would have been counted had the resources been applied.

The higher the degree of confidence the user wishes, the wider the range of possible population numbers. For example, based upon the results from the 1995 Test Census, for Oakland, CA , the true census count for a large tract with a reported total population of 10,000 would really be, with a 95% certainty, anywhere from 9,524 to 10,476, a range of plus or minus 4.8%. [Chart CA 2.2] For a typical census block of a reported population of 100 persons, the 95% confidence level means that the population could range from 65 to 135.

These ranges might seem high, they certainly do to me, but there is an important factor which affects them. The Bureau cutoff the mail follow-up at about 60% for the Test Census. The Bureau has mentioned its commitment to implementing a 90% response rate as the threshold for triggering the sampling and estimation process. Even if this is done at the tract level, the type of error shown in the Test Census will still be evident, although to a lesser degree.

As I stated about the redistricting process, the atomic level of the census block is the building block for every district. It is the starting place. If the reliability of the population counts at this lowest level is not extremely high, the results for *every* other level of census and political geography will be suspect.

It might seem that in the grand scheme of things, it wouldn't really matter whether the population for a mid-sized city with a reported population of 100,000 was really plus or minus 4.9%, or even plus or minus .49%. However, given the process of sampling and estimation used by the Bureau for the Test Census, if the range is even .49% at the higher level the range of "true" population for the atomic level of the census block will still be several times that. The methods used in the Test Census do nothing to eliminate the problems of differential accuracy at the lowest levels of geography.

Unfortunately, the court system does not operate on the grand scheme of things. The case law of these United States is quite clear that the population of each district needs to be equal, if this is practicable. In reality, it is always practicable, and thus always necessary, to create districts with equal population. However, *equal means equal*, not plus or minus 2.35%.

As the user is to ascribe their own confidence level to the results of the sampling, a time-honored and academically honest tradition, which number will the redistricting staff use? Even if the range for a census block is from 90 to 110 for a reported total of 100, should the block be factored in at 90 if it is an inner city block but 110 if it is in a suburban community? Such uncertainties will not survive strict judicial scrutiny. However, they will certainly have their day in court.

In summary, the use of these sampling and estimation techniques fails to meet the confidence levels required for either drawing districts or apportioning seats under the Constitution.

Given the court-sanctioned zero-tolerance requirements for population deviation and the small population differences which might affect the allocation of seats in the U.S. House using the apportionment formula, this would most certainly lead to more complex litigation in both areas.

The problems of differential, or, distributive accuracy for any numbers generated by the use of survey techniques will not be mitigated, let alone eliminated by this process. The use of sampling and estimation do not provide a feasible alternative to the "actual Enumeration" called for in the Constitution.

We all need to know, especially for the fair division of government funds and political power, not just *how many* persons are in these United States but also *where they live.*