WEIGHTING THE 1990 CENSUS SAMPLE

Eric Schindler, Richard Griffin, and Carolyn Swan, Bureau of the Census[1]

KEY WORDS: RAKING, SAMPLE WEIGHTING

INTRODUCTION

During the 1990 Decennial Census, a systematic sample of housing units was used to obtain housing, education, ancestry, employment, and income data not collected from the entire population. So-called long and short forms were used to collect the sample and non-sample populations, respectively. The designated sampling rates were 1-in-2 for small governmental units, 1-in-6 six in most areas, and 1-in-8 in highly urbanized areas. 54,000 weighting areas, roughly a Census tract or block numbering area, were formed. Separate sets of weights were assigned to persons and households by univariate iterative proportional fitting (raking) in four dimensions. Weighted estimates for larger race/Hispanic origin groups were controlled to their 100 percent counts. Smaller groups were collapsed with other groups to reduce the variance of the weighted sample estimates. This created a difference between the 100 percent counts and the weighted sample estimates. For the American Indian and the Hispanic origin populations operational difficulties made these differences larger than desirable. This paper describes the weighting process and some of the problems encountered and proposes areas for future investigation.

SAMPLE DESIGN

The United States is divided into about 60,000 tracts or block numbering areas (BNA). Each tract or BNA was divided into collection block groups using the first digit of the block number. Sampling rates were assigned at the collection block group level. Adequate sample for small governmental units - counties, incorporated places, and Minor Civil Divisions (MCDs) such as towns in New England - was assured by sampling those collection block groups that intersected governmental units with fewer than 2500 persons at a 1-in-2 rate (18 million persons). Other collection block groups in List/Enumerate (L/E) areas - sparsely populated areas, mostly in the West, where enumerators created the address listings and enumerated all households - or in tracts or BNAs with fewer than 2000 housing units were sampled at a 1-in-6 rate (126 million). The collection block groups in larger non-L/E tracts or BNAs were sampled at a 1-in-8 rate (105 million). Native American areas were sampled like all other areas. Hawaii has no incorporated places, so "Census Designated Places" were defined for use as governmental units. A 1-in-6 sampling rate was used for all persons living in group quarters and in Puerto Rico. For more information about the sample design and its development, see Navarro and Griffin (1990).

IMPLEMENTATION AND EDITING

Several problems in the implementation of the sample design affected the weighting process. No sample data were collected for approximately 10% of the occupied sampled households. These households were "converted" from sample forms, also known as long forms, to regular Census forms, or short forms. However, nonrespondents were not uniformly distributed by race and Hispanic origin so the weighting process was distorted. Enumerators in L/E areas often violated Census procedures by arranging their listings of housing units so as to select a disproportionate number of housing units which were either vacant or occupied by few persons. A sample tolerance check was implemented and areas with severely distorted samples were resampled. See Swan (1992). The long form had an open ended question on ancestry, and enumerators identified a higher proportion of persons as American Indians than is consistent with the 100% count of American Indians.

More extensive editing was used for sampled persons than for the non-sampled population. Some 40,000 sample persons (260,000 weighted persons) who were originally listed as being of Hispanic origin and used as such in the weighting procedure were recoded as being of non-Hispanic origin for sample estimates.

---

[1]This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau. Special thanks go to Jess Thompson, Emmet Spiers, and Donald Dalzell of the Census Bureau's Decennial Operations Division for their excellent and timely computer systems work.

SAMPLE AUGMENTATION

The first step of the weighting process, sample "augmentation", assured minimum sample sizes in all collection block groups. In block groups with very little sample, a few short form occupied housing units, vacant housing units, or group quarters persons, as needed, were randomly selected and sample data were imputed for them by a hot-deck procedure. For collection block groups in 1-in-2 areas with at least six occupied or twelve vacant housing units, the sample was augmented to at least 1-in-12. For collection block groups in 1-in-6 or 1-in-8 areas with at least twelve occupied or thirty vacant housing units or for group quarters with at least thirty persons, the sample was augmented to at least 1-in-30. Originally, the specifications allowed augmentation for occupied housing unit minima of 12 and 30 respectively, but these were lowered to 6 and 12 to allow augmentation of one household in Yellowstone Park County, Montana which originally had no sampled households and to decrease the probability of other places having no sampled households. Overall, there was very little augmentation, and most of what did occur was for group quarters persons.

INITIAL WEIGHTS

For each collection block group, four sets of initial weights equal to the relevant 100% count divided by the observed sample were calculated. The four sets were for housing unit persons, group quarters persons, occupied housing units, and vacant housing units. Since about 10% of the sample was not collected and was dropped by converting the long forms to short forms, the initial weights were generally higher than 2, 6, and 8. The initial weights provided a weighting class adjustment for nonresponse. However, they did not differentiate between differing nonresponse rates for different race or Hispanic origin groups.

FORMATION OF WEIGHTING AREAS

Tabulation block groups (TBGs) were defined as all blocks with the same first digit, in the same MCD, place, tract or BNA, and county. The TBGs in the same MCD, place, tract, and county were defined to be the initial weighting areas. Many of the initial weighting areas were too small to permit the weighting process, so collapsing was used to achieve at least 400 sample persons per weighting area. Collapsing occurred first within MCD within place, then within tract within place, then within place, then within MCD, then within sampling rate type (1-in-2, 1-in-6, 1-in-8 or mixed), and finally to the nearest tract or BNA. Weighting areas always stayed within county even if the sample was smaller than 400 persons. A total of about 60,000 weighting areas were formed.

WEIGHTING MATRIX FORMATION

Separate but similar weighting occurred for all persons and for occupied housing units. Iterative proportional fitting, or raking ratio estimation, in four dimensions was used. For person weighting the dimensions and their sizes were:

1: Race by Hispanic origin by age by sex (5x2x9x2)
2: Family type by household size (17)
3: Householder/Non-Householder (2)
4: 1-in-2/not 1-in-2 (2)

where family type was for families with children, families without children, households of unrelated persons, and group quarters. The householder variable was incorporated to make the sample estimates for householders closer to the sample estimates for households. For household weighting the dimensions and their sizes were:

1: Race by Hispanic origin of householder by owner/renter by value/rent (5x2x2x10)
2: Family type by household size (16)
3: Single unit building/2-9 units/10 or more units (3)
4: 1-in-2/not 1-in-2 (2)

where group quarters are not in the family type variable for housing units. The units in building variable has been shown to be an indicator of income. (See Swan, 1990.)

For each weighting area, four dimensional weighting matrices for persons and for occupied housing units were formed using the Census counts as the marginals and the initial weighted sample counts in the interior cells. Initially each matrix had over 10,000 interior cells. Since many cells and even some of the marginals were empty or near-empty, it was necessary to collapse the matrices in order to control the variance. Within each dimension categories were collapsed if the following conditions were not met:

1: The unweighted marginal sample count of each category had to be at least 10 for persons and 4 for housing units. This condition controls variance by not allowing excessive weights.
2: The uninflated marginal 100 percent count for a race by Hispanic origin category had to be at

least 150. This condition controls a downward bias for the race or Hispanic origin estimates which would occur if undersampled race or Hispanic origin groups were collapsed while correctly sampled or oversampled groups were allowed to remain.

3: The ratio of the marginal 100 percent count over the marginal inflated sample count could not exceed 4. This was sometimes reduced as is described later.

Collapsing started with the family type by household size variable. Groups not satisfying the conditions were collapsed with the closest group as predetermined by a set of scale values.

Race collapsing was important, because the weighted sample estimates of categories which are collapsed are not equal to the 100 percent counts. For each weighting area as many as five collapse patterns were developed to minimize the distortion. One collapsed all races not meeting the criteria together to see if the criteria were met. One collapsed races not meeting the criteria with other races with the closest observed sampling rates until the criteria were met for all surviving groups. The others collapsed all races not meeting the criteria with each of the races meeting the criteria. The specifications called for the estimation of two "bias" terms for each collapse pattern. The first was the sum of the absolute values of the differences between the 100 percent count and the weighted sample estimate for each race using the initial weights. The second included only that portion of each above which exceeded 1% of the 100 percent count. This factor gave emphasis to obtaining good sample estimates for the smaller races. The collapse pattern with the smallest sum of the two factors would be selected. See Schindler (1990). However, improved results were obtained by using only the second factor, giving additional emphasis to the small groups.

For the remaining race categories, Hispanic origin was collapsed if the criteria were not met for both the Hispanic and the non-Hispanic portions of the race group. For the remaining race/Hispanic origin categories, the 18 age/sex categories or 20 rent/value categories (which were all in the same dimension of the matrix) were collapsed according to prescribed scale values.

Collapsing of the householder/non-householder or units in building and sampling rate dimensions occurred if the criteria were not met. Units in building categories were collapsed first with the nearest or nearest/smallest category.

RAKING AND WEIGHT ASSIGNMENT

The raking procedure adjusted the weights of the family type by household size dimension first, then the sampling rate dimension, then the householder or units in building dimension, and finally the race by Hispanic origin by age/sex or race by Hispanic origin by rent/value dimension. Two iterations were performed and a controlled rounding was used at the eighth and last adjustment to produce integer values for the weighted sample estimates in each remaining cell. Next, to control excessive weights, the largest ratios of the final weighted cell estimates divided by the initial weighted cell estimates (for persons or for occupied housing units) were multiplied by the largest (person or group quarters or occupied housing unit) initial weights, separately for collection blocks groups sampled at 1-in-2 in which the observed rate was greater than 20% and for all other collection block groups. If these largest possible weights were greater than 12 in the 1-in-2 areas or greater than 48 elsewhere, additional collapsing of the initial weighting matrix was needed. Collapsing was forced by reducing the maximum allowable ratio of the marginal 100% count divided by the marginal initial weighted sample estimate from 4.0 to 3.0 and then to 2.6, 2.2, 1.8,... as necessary to reduce the size of the matrix. Raking and, if necessary, additional collapsing would be repeated until the maximum achievable weights were less than 12 and 48. The initial weight of each person or housing unit was then multiplied by the ratio of his/its final weighted cell estimate divided by his/its initial weighted cell estimate and then control rounded to obtain final weights.

OPERATIONS

Greater detail on the weighting procedures from augmentation through weight assignment can be found in the computer specifications by Griffin, Swan, and Schindler (1990) which were completed at the end of 1990. During the spring of 1991, specifications for review listings were completed. Computer systems work by the Census Bureau's Decennial Operations Division was completed during the summer and fall of 1991. Weighting listings were reviewed first to assure that the specifications were properly implemented and then to assure that the results were appropriate during the fall of 1991. Weights for all fifty states and the District of Columbia were cleared for sample data estimation in January of 1992.

No operation as complex as estimating consistent and appropriate sample weights for 40 million persons in 17 million housing units is completed flawlessly. Good working relations between the systems staff and the statistical staff made the identification and resolution of the several minor problems which occurred as fast and painless as possible.

There were, however, several problems caused by inadequacies in the implementation of the Census sample which could not be resolved by changes to the weighting system. They were identified at the state level by comparing the 100 percent counts of population with the weighted sample estimates, with or without the editing of the race and Hispanic origin fields. Table 1 shows these estimates and the percent change from the 100 percent counts at the national level.

Table 1: Population Estimates by Race (000s)

	100% Counts	Weighted Sample	Edited Sample
Total	248,710	248,710	248,710
White	199,686	199,791 0.05%	199,830 0.07%
Black	29,986	29,920 -0.22%	29,927 -0.20%
AmInd	1,959	2,010 2.58%	2,016 2.88%
Asian	7,274	7,217 -0.78%	7,227 -0.65%
Other	9,805	9,771 -0.34%	9,710 -0.97%
Hisp	22,354	22,161 -0.86%	21,899 -2.03%

No concern was raised over the differences for Whites or Blacks. Most persons listing their race as "Other" are of Hispanic origin and they share the same problems. At least a portion of the differences for Asians may be linked to American Indians. Possible causes for the differences for American Indians and persons of Hispanic origin have been identified. No acceptable solutions using the weighting system were identified, but the problems will be studied over the next several years to try to avoid similar differences in 2000.

AMERICAN INDIANS

Weighted sample estimates account for 50,000 more American Indians than the 100 percent counts. Most of this difference arises in those states east of the Rocky Mountains without large reservations. In general, American Indians living on reservations were undersampled but the raking procedure forced the weighted sample estimates to equal the 100 percent counts. On the other hand, American Indians not living on reservations were generally oversampled. Since there were fewer than 150 of these American Indians in most weighting areas, they were generally collapsed with another race and their weighted sample estimate was not raked back down to the 100 percent count. Weights were recalculated for several states without the population requirement of 150 persons for the American Indian group, requiring that only the less rigorous 10 sample person criterion be met. This "quick fix" reduced the difference by only 20%, so it was not used.

Counts for individual tribes were examined. The differences between the weighted sample estimates and the 100 percent counts for Cherokees, the largest, best-organized tribe in these areas, is most of the total difference. Much of the oversample seemed to be in the collection of long forms by enumerators. The long form mailback rate for American Indians was slightly lower than the short form mailback rate, consistent with other races. The long form enumerator-filled rate, however, was much higher than required to compensate for the slightly lower mailback rate.

A sample of long and short forms, mailback and enumerator-filled forms for Georgia and Tennessee, two seemingly typical states for this problem, was examined. (Schindler, 1992.)

In addition to the weighted sample counts overestimating the 100% counts by about 15%, the 100% counts appeared to be overestimates of the actual American Indian population by about 8%. Four major sources of error were found:

1. Asian Indians, East Indians, and West Indians enumerated as American Indians increased the 100% counts and weighted sample estimates for American Indians by about 1%.
2. Canadian and Latin American Indians enumerated as American Indians increased the 100% count by about .5% and the weighted sample estimate by about 2%.
3. Blacks miscoded as American Indians increased the 100% count by about 1% and the weighted sample estimate by about 2%.

4. Erasure rates on forms filled out by enumerators far exceeded those on forms mailed back. Some of these changes were legitimate, but some enumerators may have created American Indians from persons who did not consider themselves to be American Indians even though some of their ancestors were American Indians, often Cherokee. Erasures exceeding the proportion found on mailback forms increased the 100% count by about 4% and the weighted sample estimate by about 7.5%.

In all, the anomalies uncovered reduced the 100% count by about 7.5%, the weighted sample estimate by about 12.5%, and the difference between the two by about 40%.

PERSONS OF HISPANIC ORIGIN

As seen in Table 1, 1.17% of the 2.03% of the difference between the weighted sample estimates and the 100 percent counts for persons of Hispanic origin was caused by additional editing of the sample file. Weights were calculated using the initial sample file to be consistent with the less stringent short form edit procedures. During the additional sample editing a substantial number of persons who had marked one of the Hispanic origin circles were deemed to be not of Hispanic origin because their answers to the ethnic origin and language spoken at home questions showed no Hispanic ties. The recoded Hispanic origin field was considered to contain better information for estimation than the initial data. This portion of the difference was created after the weighting was performed. There was no appropriate procedure to adjust the weighting to reduce this difference.

The remaining 0.86% of the difference between the weighted sample estimates and the 100 percent counts for persons of Hispanic origin has been linked with two sources. The first was caused by the fact that, after the initial editing, answers to the Hispanic origin question had to be allocated to about 12% of the short forms but to only 4% of the long forms. Persons of Hispanic origin answered the Hispanic origin question at a higher rate than the remaining population, so the hot-deck donor pool contains an overabundance of Hispanics. Therefore, too many persons well allocated as being of Hispanic origin for both the long and short forms. The higher allocation rate for short forms meant that Hispanic origin was erroneously allocated to more persons for the 100 percent counts than for the weighted sample estimates. To the extent that the 193,000 person difference between the weighted sample estimates and the 100% counts is caused by this allocation bias, the weighted sample estimates may be closer to the "truth" than the 100% counts.

The second possible source for the difference of 0.86% between the weighted sample estimates and the 100% counts for persons of Hispanic origin was in the weighting system itself. The observed sampling rates for Hispanics were lower than the overall observed sampling rates, most likely caused by higher nonresponse rates for persons of Hispanic origin. However, there may have been some interviewer effect in some areas if enumerators tried to steer the sample and its longer form to English-speaking households. Since many of the remaining race groups in many weighting areas had less than 10 sample persons or 150 persons of Hispanic origin, many of the Hispanic origin groups were collapsed with the non-Hispanic origin group. The raking procedure forced the weights for the entire collapsed race group to be equal to the 100% count. However, the less well sampled Hispanic origin groups were assigned weights less than the 100% counts and the better sampled non-Hispanic origin groups were assigned weights more than the 100% counts. An attempt was made to reduce this collapsing bias. The number of collapse patterns considered was increased by creating additional patterns where the White and Other races were forcibly collapsed together. Since most Hispanics are in these two races, it was hoped that the Hispanic origin group would not have to be collapsed as often. Also, the formula used for the selection of collapse pattern was changed to include the difference for persons of Hispanic origin. Forcing the use of collapse patterns where Whites and Others were collapsed produced only a 20% reduction in the differences for Hispanics, but added to the differences for various race groups, particularly American Indians. To the extent that the 193,000 person difference between the weighted sample estimates and the 100% counts is caused by this collapsing bias, the 100% counts may be closer to the "truth" than the weighted sample estimates.

ON TO 2000

Given the operational difficulties to be expected with any program the size of the Decennial Census, the weighting of the 1990 Census sample was a success. Statistically sound methodologies were used. Potential causes for the most serious discrepancies were identified. Alternative procedures within the basic methodological framework were used to try to address the problems. And the process was completed on schedule.

This basic satisfaction does not indicate complacency. There are several possible improvements to test in preparation for the Decennial Census for the year 2000. These will be grouped into at least four empirical studies over the next several years. The first will explore several alternatives within the iterative proportional fitting framework. Two questions will be addressed:

1. Are the discrepancies between 100% counts and weighted sample estimates decreased and the sample estimates improved if the initial weights within the collection block groups are calculated (1) separately for the larger race/origin groups or (2) by logistic regression or (3) by the first stage of a two-stage raking procedure?
2. Are sample estimates improved if, instead of converting nonresponding sample households to nonsample households, sample data are allocated to these households based on the available sample data? This question took on added importance in 1990 when the rate of noncooperation for long forms tripled to 10% compared to 3.5% in 1980.

The second empirical study will explore alternatives to the iterative proportional fitting methodology. Six similar alternatives minimizing different metrics will be compared:

1. Univariate iterative proportional fitting, the current system
2. Multivariate iterative proportional fitting in which the weights of persons and households are assigned simultaneously with all persons in a household receiving the same weight as the household
3. Univariate maximum likelihood
4. Univariate minimum chi-squared
5. Univariate generalized least squares
6. Multivariate generalized least squares for both the persons and housing units simultaneously.

The third empirical study will investigate whether empirical Bayes smoothing of the weights could reduce the sampling error for small area estimates. The fourth study will investigate the existence and structure of nonignorable response mechanisms and the development of appropriate nonresponse adjustment models in place of the hot-deck procedure as a means of reducing error estimates.

REFERENCES

GRIFFIN, Richard, SWAN, Carolyn, and SCHINDLER, Eric, 1990, 1990 Decennial Census Sample Weighting, "STSD 1990 Decennial Census Memorandum Series, W-35 and various Addenda.

NAVARRO, Alfredo and GRIFFIN, Richard, 1990, Sample Design for the 1990 Decennial Census, Proceedings of the Section for Survey Research Methods", 1990 Meeting of the American Statistical Association

SCHINDLER, Eric, 1990, Race Collapse Procedure, "STSD 1990 Decennial Census Memorandum Series", W-46

SCHINDLER, Eric, 1992, Microfilm Study for American Indians in Georgia and Tennessee, "STSD 1990 Decennial Census Memorandum Series", W-63

SWAN, Carolyn, 1992, The 1990 Census Sample Tolerance Check: Evaluation Results, Proceedings of the Section for Survey Research Methods", 1992 Meeting of the American Statistical Association

SWAN, Carolyn, 1990, Poststratification Designs for Approximating the Effect of Dimension Variables under Iterative Proportional Fitting, "STSD 1990 Decennial Census Memorandum Series", W-47.