EFFORTS TO IMPROVE THE RELIABILITY OF MINORITY ESTIMATES IN THE REDESIGN OF THE CURRENT POPULATION SURVEY

Deborah Lorah, Rajendra Singh, and Robert Tegels, U.S. Bureau of the Census

I. INTRODUCTION--THE CPS DESIGN

The Current Population Survey (CPS) is a complex multistage stratified sample survey of households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The survey produces the monthly estimates of the labor force characteristics that provide detailed information on the economic status and activities of the civilian noninstitutional population. In 1978, the Congress established the National Commission on Employment and Unemployment Statistics (NCEUS) to study the meaning of our labor force data system and to make whatever recommendation may be necessary to improve the system's accuracy and relevance to current conditions. After over one year's extensive work, the Commission submitted a report [1] containing altogether about 100 recommendations. One of the recommendations was that the CPS should yield more reliable national data for minority (black and Spanish origin) groups.

Every 10 years the CPS sample is redesigned to to take advantage of the most recent decennial census data and improved sampling and operational techniques. We are currently designing a new CPS sample which is scheduled to be in place and functioning in January 1984. This paper discusses some highlights of our research toward designing this new sample to improve the reliability of minority estimates. Specifically, we will be concerned with reducing the variance on monthly estimates of black and Spanish unemployed at the national level in such a way that the established reliability of estimates of total unemployment is maintained.

The present CPS design consists of about 77,000 designated housing units nationwide. Before the sample was selected, the entire country was divided into Primary Sampling Units (PSUs), each typically consisting of a county, group of counties, or independent city. The PSUs were then grouped into strata, and either one or two PSUs was drawn from each stratum with probability proportionate to its population size as of the 1970 Census. A sample PSU selected from a stratum containing more than one PSU is said to be nonself-representing (NSR). Some PSUs were so large that they were placed into strata all by themselves. They are in sample with certainty and are said to be self-representing (SR). Within each sample PSU a sample of Ultimate Sampling Units (USUs) is selected using systematic sampling. These USUs are clusters of housing units, usually geographically contiguous and having an expected size of four. A detailed description of the CPS design is given in [2].

It can be seen from the brief description presented here that the variance of a CPS estimate can be divided into two components: a between-PSU component, arising because only a sample of all PSUs is selected, and a within-PSU component, arising because only a sample of USUs is selected from each sample PSU. (Note that an SR PSU has no between-PSU component of variance.) The methods discussed in this paper are aimed at reducing the within-PSU component of variance for minority estimates. The two major ideas to be discussed are: 1) stratification by race within the PSU, either at the housing unit or Enumeration District level, and 2) changing the size of the USU.

It must be stressed that the ideas discussed in this paper will not automatically be incorporated into the new CPS design. Costs and operational feasibility must also be considered. At the present time, stratifying and oversampling within PSUs does not seem likely.

II. STRATIFICATION WITHIN THE PSU BASED ON MINORITY STATUS

One way to improve the reliability of minority estimates would be to somehow identify all of the minority-occupied housing units in the PSU prior to sampling, group them into a separate "minority" stratum, and sample them at a higher rate than the non-minority housing units.

Since part of our sampling frame will consist of addresses from the 1980 Census file, we could ascertain prior to sampling which housing units were occupied by blacks or persons of Spanish origin at the time of the census. Unfortunately, there is no guarantee that a housing unit occupied by a black family, for example, in 1980 will remain occupied by blacks over the next decade. This is important because all of the CPS sample to be used for the next 10 years will be selected at the same time. If the number of minority households in the "minority" stratum decreases over time, any gains made in reliability due to oversampling in this stratum may vanish.

Thus, before we stratify at the housing unit level, we should examine the racial stability of the housing inventory over time. We should also try to achieve the desired reliability with a minimum sample size and minimum effect on the reliability of total unemployment. We should examine the effect on reliability when the racial composition of the "minority" and "non-minority" strata changes over time. Finally, if stratification at the housing unit level proves infeasible, we should consider stratifying at a larger geographic level within the PSU, such as at the Enumeration District (ED) level.

A. Racial Distribution of the Housing Inventory

The CPS sampling frame consists primarily of addresses from the most recent census (old construction). It is supplemented by another frame representing housing units built since the census (new construction). Changes in the racial composition of both of these frames over time will affect the variance of minority estimates.

Old Construction--The data files for two Census Bureau surveys, the Survey of Income and Education (SIE) and the 1976 Survey of Registration and Voting (RAV) were used to determine the racial status of individual sample housing units at two points in time.

SIE data for the United States, the four census regions, metropolitan, and nonmetropolitan areas, were tabulated showing the percentage of 1970 nonwhite-occupied housing units still occupied by nonwhites in 1976 and the percentage of 1970 white-occupied units still occupied by whites in 1976. According to these tabulations, at the national level, 89.3 percent of all housing units occupied by nonwhites in 1970 were

still occupied by nonwhites in 1976. At the regional level 73.6 percent (west) to 93.8 percent (south) of the 1970 nonwhite housing units were still occupied by nonwhites in 1976. The percentages in metropolitan and nonmetropolitan areas were little different from each other or the national percentage of 89.3. Detailed tabulations and analysis of the SIE data can be found in [3].

Similar tabulations were made of Spanish/non-Spanish turnover and white/black/other nonwhite turnover using RAV data for the following geographic areas (voting jurisdictions):

A: Ten counties in Arizona, New Mexico, Colorado, and California.
B: Two counties in Arizona.
C: Pinal County, Arizona.
D: Parts of Florida and New York City.
E: Alabama, Georgia, Louisiana, Mississippi, South Carolina, Virginia, and part of North Carolina.
F: Robeson County, North Carolina.

Of all housing units occupied by Spanish heads in 1970, the proportion that was occupied by Spanish heads in 1976 was 57.5 percent in geographic area A, 59.6 percent in Area B, 72.5 percent in Area C, and 65.7 percent in area D. Of households occupied by non-Spanish heads in 1970, 93.2 percent were occupied by non-Spanish in 1976 in area A, 94.0 percent in area B, 82.2 percent in area C, and 74.8 percent in area D.

For housing units occupied by whites in 1970, 93.9 percent were still occupied by whites in 1976 in area E, 90.5 percent in area F, and 67.6 percent in area D. For housing units occupied by blacks in 1970, 95.8 percent were occupied by blacks in 1976 in area E, 89.7 percent in area F, and 84.9 percent in area D. For households occupied by other nonwhites in 1970, 39.8 percent were still occupied by other nonwhites in 1976 in area E, 83.1 percent in area F, and 28.4 percent in area D. For details, see [4].

New Construction--Data from the Annual Housing Survey (AHS) show that at the national level blacks were underrepresented in new construction housing while whites were overrepresented. Blacks occupied 10.6 percent of all housing units but only 7.1 percent of new construction; whites occupied 87.8 percent of all housing units and 91.2 percent of the new construction. Similar differentials were observed at the regional level and for metropolitan/nonmetropolitan breakdowns. For details, see [5].

Conclusions--According to the data from SIE and RAV, it would seem feasible to stratify 1980 black and nonblack-occupied housing units into a minority and nonminority stratum prior to sampling, as about 90 percent of these black-occupied units could be expected to remain black during the life of the survey. New construction, as it would be expected to produce small and underrepresentative numbers of black-occupied units, would be put into the nonminority stratum. Stratification at the housing unit level does not seem as feasible for improving the reliability of Spanish estimates, as almost half of the Spanish-occupied units may change to non-Spanish-occupied in later years.

B. Stratification at the Housing Unit Level

1. Choosing Sample Sizes.--Consider a PSU where a certain reliability must be maintained on estimates of total unemployed. At the same time we want to reduce by a certain factor the variance on the number of unemployed persons in a specified minority.

Let N = total population 16+;
X = the variable total unemployed persons;
Y = the variable unemployed minority persons;
n = sample size needed to meet the reliability requirement for total unemployed if a simple random sample is drawn;
N_B = total minority population 16+;
σ_X^2 or σ_Y^2 = population variance for the variable X or Y.

Instead of selecting a simple random sample, we could subdivide the population in the PSU into two strata and select samples of size n_1 and n_2 within each stratum such that for this stratified design the reliability requirement for X is still met, but the variance on the estimated level of Y is some predetermined proportion f of what it would have been if a random sample of size n had been chosen.

Let N_1, N_2, N_{1B}, N_{2B}, σ_{X1}^2, σ_{X2}^2, σ_{Y1}^2, and σ_{Y2}^2 be defined analogous to N, N_B, σ_X^2, and σ_Y^2.

Then the desired stratum sample sizes are given by:

$$n_2 = \left[\frac{N_2 \sigma_{X2}^2 N_{1B} \sigma_{Y1}^2 - N_1 \sigma_{X1}^2 N_{2B} \sigma_{Y2}^2}{N \sigma_X^2 N_{1B} \sigma_{Y1}^2 - f N_B \sigma_Y^2 N_1 \sigma_{X1}^2} \right] \frac{n N_2}{N}$$

$$n_1 = \frac{N_1^2 \sigma_{X1}^2}{\frac{N^2}{n} \sigma_X^2 - \frac{1}{n_2} N_2^2 \sigma_{X2}^2}$$

The new sample size is $n' = n_1 + n_2$.

Now suppose we are able to stratify by race so that stratum one contains only minority persons and stratum two contains everyone else.

Then,

$$n_1 = \frac{1}{f} \frac{n N_1}{N} \text{ and } n_2 = \left[\frac{N_2 \sigma_{X2}^2}{N \sigma_X^2 - f N_1 \sigma_{Y_1}^2} \right] \frac{n N_2}{N}.$$

To illustrate this method, sample sizes for a sample stratified by race (assuming the first stratum contained only minority units) were calculated for the Indianapolis and Atlanta SMSAs, and for the District of Columbia using 1970 census data. They are presented in table 1. The initial sample sizes n were calculated using current CPS within-PSU sampling intervals.

In order to reduce the variance by a half under simple random sampling, n would have to be doubled. It can be seen from the table that for a stratified sample the reduction in variance can be achieved with a sample size less than 2n, and possibly less than n.

2. Deterioration of Reliability Over Time.--Since the turnover rate of Spanish households in individual housing units seems to be rather high, we estimated the expected deterioration in reliability for a stratified sample due to changing strata composition. We also compared the variances for the stratified sample and a simple random sample.

Suppose that all the 1980 housing units in a sample PSU are stratified into a Spanish and a non-Spanish stratum:

Stratum I

(1) Contains all Spanish housing units in 1980 (time o).
(2) A proportion p_1 of the units are Spanish at a given later date (time t).
(3) The stratum is sampled at a sampling interval r_1 at both points in time.
(4) Contains a proportion p of all the housing units in 1980 and the same proportion at the later date.

Stratum II

(1) Contains no Spanish housing units in 1980.
(2) A proportion p_2 of units are Spanish housing units at the later date.
(3) The stratum is sampled at a sampling interval r_2 at both points in time.
(4) Contains a proportion 1-p of all the housing units in 1980 and the same proportion at the later date.

We further assume here that the proportion of the population in Stratum I equals the proportion of the housing units in Stratum I, and the proportion of the population that is Spanish equals the proportion of the housing units that is Spanish.

Let σ_1^2 be the population variance over Stratum I, σ_2^2 the population variance over Stratum II, and σ^2 the population variance over the whole area without regard to strata. We assume that σ^2, σ_1^2, and σ_2^2 remain constant over time.

To further simplify the formulae, let

$$p_1 = u\,p_2$$
$$1-p = v\,p$$
$$r_1 = k\,r_2$$
$$\sigma_1^2 = w\,\sigma_2^2$$

Then the ratio of the variance of the estimated level of Spanish unemployed for the stratified sample at time t, $VAR_t(SS)$, to the variance for the stratified sample at time o, $VAR_o(SS)$, is given by:

$$\frac{VAR_t(SS)}{VAR_o(SS)} = P_2\left(u + \frac{v}{kw}\right)$$

Suppose that a simple random sample of the same size was drawn instead, without regard to strata. Then the ratio of the variance for the stratified sample at time t to the variance for the simple random sample at time t, $VAR_t(SRS)$, is

$$\frac{VAR_t(SS)}{VAR_t(SRS)} = \frac{(k\,u\,w + v)(1 + k\,v)\,\sigma_2^2}{k(v+1)(u+v)\,\sigma^2}$$

Data from RAV for areas A, B, and D were used to simulate a stratified design under the assumptions that all Spanish occupied housing units had been placed in Stratum I and $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Then the above variance ratios were calculated for several values of k.

It can be seen from table 2 that in most cases the variance for the stratified sample will increase over time. The exception was area (B) in Arizona where reductions did occur for k=.9 and k=.8. For most values of k, however, the 1976 variance for a stratified sample was still lower than the variance for a simple random sample, as shown in table 3. Only for k=.2 and k=.1 did increases occur. Thus, it looks like stratification would have been feasible, i.e., better than a simple random sample, in these three areas, even though the Spanish composition changed considerably over the 1970-1976 period.

C. Stratification at the Enumeration District Level.

We also considered stratifying within PSUs at the Enumeration District (ED) level. (EDs are PSU subdivisions typically defined to contain an expected 350 housing units.) Since the Spanish population seemed to be more concentrated in specific areas of the country than blacks, we decided to test an ED stratification plan on 24 SMSAs that had Spanish populations of over 50,000 persons at the time of the 1970 census. Again, we were interested in the reductions in variance possible when a differential stratified sample is selected rather than a simple random one. The sampling approach taken is similar to the one used by Waksberg in [6].

Sampling and Stratification Methodology

Consider two sampling plans:

A: A simple random sample of size n is selected without regard to strata.

B: The PSU is stratified and a simple random sample is selected from each stratum using sampling intervals r_1 and r_2 such that the total sample size is n.

Define:

v = total population in stratum 2 divided by total population in stratum 1.

u = proportion Spanish 16+ in stratum 1 divided by proportion Spanish 16+ in stratum 2.

$k = r_1/r_2$.

$\sigma_1^2(Y)$ or $\sigma_2^2(Y)$ = population variance for Spanish unemployment in stratum 1 or stratum 2.

$\sigma_1^2(X)$ or $\sigma_2^2(X)$ = population variance for total unemployment in stratum 1 or stratum 2.

$\sigma^2(X)$ or $\sigma^2(Y)$ = population variance for total unemployment or Spanish unemployment over the entire PSU.

Then the ratio of the variances of Plan B to Plan A (i.e., the design effect for differential sampling) for Spanish unemployed is:

$$\frac{VAR_Y(B)}{VAR_Y(A)} = \frac{\sigma_2^2(Y)\,[k\,u\,\sigma_1^2(Y)/\sigma_2^2(Y) + v]\,(1 + k\,v)}{\sigma^2(Y)\,k\,(1+v)\,(u+v)}$$

This ratio is minimized for $k = 1/\sqrt{u\,\sigma_1^2(Y)/\sigma_2^2(Y)}$.

The ratio of variances for total unemployment is given by:

$$\frac{VAR_X(B)}{VAR_X(A)} = \frac{\sigma_2^2(X)\,[k\,\sigma_1^2(X)/\sigma_2^2(X) + v]\,(1 + k\,v)}{\sigma^2(X)\,k\,(1+v)^2}$$

For each SMSA, successive pairs of strata were created using data from the 1970 census. The EDs were first sorted from high to low on the basis of proportion of persons of Spanish origin aged 16+. Then, for the i^{th} pair of strata

stratum 1 = {1st ED, 2nd ED, ... , i^{th} ED}
stratum 2 = {all remaining EDs}

For each pair of strata formed, $VAR_X(B)/VAR_X(A)$ and $VAR_Y(B)/VAR_Y(A)$ were computed using $k = 1/\sqrt{u\,\sigma_1^2(Y)/\sigma_2^2(Y)}$.

For each SMSA, the variance ratios were compared for each pair of strata. Lack of time has prevented a full analysis of the resulting data. Some preliminary results, however, are described below.

The biggest reduction in the variance for Spanish unemployed due to differential stratified sampling, almost 70 percent, occurred in Newark, where the Spanish population constituted 1.5 percent of the total population in 1970. The variance on total unemployment, however, increased almost 300 percent. In other SMSAs the reductions for Spanish were more modest, but the corresponding increase in variance for total unemployed was also smaller. In San Jose, for example, the Spanish variance decreased by 22.2 percent while total unemployment experienced only an 11.4 percent increase. In many SMSAs the decrease in variance for Spanish was accompanied by an increase for total of about the same magnitude. An exception was McAllen, Texas, where 73.5 percent of the total population was Spanish. In this SMSA the Spanish variance decreased 3 percent and the total variance also decreased very slightly.

Looking at it another way, stratifying Newark, Chicago, Jersey City, and Detroit resulted in a 20 percent or more decrease in the variance for Spanish unemployed while allowing only a 5 percent increase in variance for total unemployed. It should be noted that in each of these SMSAs the proportion Spanish was very low. In El Paso, half of which was Spanish in 1970, only a 9 percent reduction for Spanish could be achieved while causing a 5 percent increase for total. The results were similar for a 10 percent increase in variance for total unemployed. Detailed tabulations are given in [7].

III. THE SIZE OF THE ULTIMATE SAMPLING UNIT

Another method of increasing reliability for minority estimates is to reduce the size of the ultimate sampling unit (USU). Currently, the USU used for CPS is a cluster of expected size four adjacent housing units. Previous studies had indicated that the variance for the major labor force items could be minimized at a fixed cost by using USUs of this size. But while this method of sampling is advantageous for the population as a whole, it results in less reliable estimates for characteristics of subgroups which tend to cluster. Since minorities who live in the same neighborhoods tend to have the same characteristics to a greater degree than do nonminorities, it was felt that minority estimates could be made more reliable if the size of the USU were reduced.

Processing.--In order to test how much gain could be made by reducing the USU size, an empirical study was conducted using three independent months of CPS data. First, a set of minority characteristics was selected for study. For each of these characteristics, an estimate and estimated variance were calculated using the standard CPS methods. Next, USUs of expected size three and two were created by randomly deleting 25 and 50 percent of the housing units in each USU from the original data file. After the housing units to be deleted were determined, estimates and variances for the same characteristics were calculated. These estimates and variances were averaged over the 3 months and the analysis was performed using the averages.

Theory.--In order to estimate the effect of reducing the size of the USU on the variance, two assumptions had to be made. First, although the CPS takes a systematic sample of USUs, it was assumed that the sample was a simple random sample. Next it was assumed the variations in the number of persons per USU had little effect on the study. There was an average of two persons per housing unit in the CPS, so that USUs of size eight, six, and four persons corresponded to USUs of size four, three, and two housing units. Under these assumptions the percent reduction in variance by changing from USUs of size $\bar{n}$ to size $\bar{n}'$ with the same total sample size is given by:

$$\left[1 - \frac{1 + (\bar{n}' - 1)\,\delta_{\bar{n}'}}{1 + (\bar{n} - 1)\,\delta_{\bar{n}}}\right] \times 100$$

where $\bar{n}$ is the average number of persons per USU

$\delta_{\bar{n}}$ is the interclass correlation for USUs of size $\bar{n}$. This is a measure of how the people in the USU tend to have similar characteristics.

An estimate of $\delta_{\bar{n}}$ is: $\hat{\delta}_{\bar{n}} = \frac{DEFF - 1}{\bar{n} - 1}$ where DEFF is the within PSU design effect.

In order to calculate DEFF, note that in a simple random sample, the variance can be calculated as

$$Var(x)_{SRS} = Var(N\tfrac{x}{N}) = N^2\ Var(\tfrac{x}{N}) = SI \cdot (1 - \tfrac{x}{N}) \cdot x$$

where SI = is the sampling interval used in selecting CPS sample

N = Total Population

$$\text{Hence, } DEFF = \frac{Var(x)}{SI \cdot x \cdot (1 - \frac{x}{N})} .$$

This formula represents the design effect for USUs of size four housing units. For the USUs of size three housing units:

$$DEFF^1 = \frac{Var(x)}{SI \cdot x \cdot (1 - \frac{4}{3}\frac{x}{N})}$$

In this formula, x and Var(x) are obtained directly from the estimates and variance program.

Similarly, for USUs of two housing units:

$$DEFF = \frac{Var(x)}{SI \cdot x \cdot (1 - \frac{2x}{N})}$$

Results.--In switching from USUs of size four housing units to three, the variance was decreased by 16 percent on the unbiased estimate of of total black population, by 12 percent for Civilian Labor Force, 6 percent for Unemployed, and 15 percent for persons Not in Labor Force. For Persons of Spanish Origin, there were decreases of 13 percent for Total Population, 11 percent for Civilian Labor Force, and 12 percent for Not in Labor Force. There was no significant decline in variance of Spanish Origin Unemployed. For the population as a whole, the variance for CLF declined by 19 percent, while there was no significant change in the variance for unemployed.

Larger declines in variance were observed in switching from USUs of size four to size two. The variance for black population declined by 36 percent, while Civilian Labor Force declined by 30 percent, Unemployed by 16 percent, and Not in Labor Force by 29 percent. For persons of Spanish Origin, the declines were 32 for Total

Population, 25 for Civilian Labor Force, and 29 for Not in Labor Force. Again, there was not a significant decline in the variance of Spanish Unemployed. For the Total, CLF declined by 38 percent and Unemployed by 9 percent. For more details, see [8].

These results apply only to the CPS unbiased estimates. The reductions in variance of the CPS ratio estimates are still to be determined. The next step would be to conduct a cost study. Decreasing the USU size would increase the costs needed to maintain present reliability on estimates for the population as a whole since USUs of size four have previously been determined to be optimal. It must be determined whether this increase in cost would be offset by the cost savings for minority estimates by using a smaller USU. Otherwise, if USUs remain size four, the sample size would have to be increased in order to improve the minority estimates. A compromise solution may be to retain USUs of size four throughout most of the country, but in areas with high minority concentrations, use a smaller USU if a smaller USU is in fact optimal for minorities. The operational difficulties of this type of sampling operation also would have to be studied.

IV. CONCLUSIONS

Stratifying the housing units based on the racial status will improve the reliability of estimates for blacks since a very high proportion of black housing units (HUs) would remain black for the life of the survey. This approach is not as encouraging for Spanish. The Spanish HUs change status at a much higher rate. However, most persons of Spanish origin are concentrated in a limited number of SMSAs and a few states. Therefore, oversampling in these SMSAs and states would be an appropriate approach to improve the Spanish estimates.

Another approach to improving the minority estimates involves stratifying the EDs within a PSU into two strata and then oversampling the stratum with the higher minority concentration. This approach may prove efficient in some SMSAs if we are willing to accept some increase in the variance for total unemployed or in the sample size. More analysis of the available data is needed, however, before we decide whether certain PSUs are good condidates for stratification and how they could be stratified.

Minority estimates could also be improved by reducing the USU size. If this alternative is adopted, the USU size could be reduced to either three or two from the current size of four.

One or more approaches could be used to improve minority estimates. However, before any final decision is made, the impact on field and processing operations and the cost must be evaluated.

FOOTNOTE

[1]The estimate must be increased by a factor of 4/3 because when calculating the estimates and variances, the weights on the sample records weren't changed to reflect the deletion of sample units.

REFERENCES

[1] National Commission on Employment and Unemployment Statistics. "Counting the Labor Force." Stock No. 052-003-00695-2, Washington, D.C. Government Printing Office, 1979.

[2] U.S. Department of Commerce, Bureau of the Census. "The Current Population Survey: Design and Methodology," by Robert Hanson. Technical Paper 40. Superintendent of Documents No. C3.212:40; January 1978.

[3] Lorah, Deborah J. "CPS Redesign: Minority Estimates--Results of Race Tabulations from the SIE File." Bureau of the Census internal memorandum for Documentation. April 28, 1980.

[4] Lorah, Deborah J. "CPS Redesign: Minority Estimates--Results of Minority Tabulations from the 1976 RAV Data File." Bureau of the Census internal memorandum for Documentation. August 20, 1980.

[5] Lorah, Deborah J. "CPS Redesign: Minority Estimates--Distribution of New Construction Housing Units by Race of Occupants." Bureau of the Census internal memorandum for Documentation. May 20, 1980.

[6] Waksberg, Joseph. "The Effect of Stratification with Differential Sampling Rates on Attributes of Subsets of the Population." Presented at the 133rd Annual Meeting of the American Statistical Association, August 1973.

[7] Lorah, Deborah J. "CPS Redesign: Minority Estimates--Preliminary Results of Stratification of EDs to Improve Reliability of Spanish Unemployment Estimates." Bureau of the Census internal memorandum for Lawrence S. Cahoon, October 1, 1981.

[8] Cahoon, Lawrence. "CPS Redesign: Minority Estimates--Change in Segment Size to Reduce the Variance of Minority Estimates." Bureau of the Census internal memorandum for Gary M. Shapiro. Prepared by Robert Tegels, January 21, 1981.

TABLE 1. Sample Sizes for Stratified and Simple Random Sampling

CITY	Proportion Nonwhite	n	n_1	n_2	n'
Indianapolis	.115	711	164	542	706
Atlanta	.207	561	234	358	592
District of Columbia	.649	900	1168	122	1290

TABLE 2. $\frac{VAR_t(SS)}{VAR_o(SS)}$

Area	p_1	p_2	p	K: .9	.8	.6	.4	.2	.1
A	.575	.068	.130	1.081	1.144	1.333	1.713	2.850	5.124
B	.596	.060	.158	.951	.996	1.129	1.395	2.195	3.794
C	.657	.252	.152	2.220	2.415	3.001	4.173	7.689	14.716

TABLE 3. $\frac{VAR_t(SS)}{VAR_t(SRS)}$

Area \ k	.9	.8	.6	.4	.2	.1
A	.958	.917	.844	.795	.841	1.080
B	.951	.904	.818	.754	.782	1.003
D	.984	.972	.961	.993	1.198	2.689