

September 13, 1995

MEMORANDUM FOR Elizabeth Martin  
Chief, Center for Survey Methods Research  
Statistical Research Division

From: Yukiko Ellis  
Center for Survey Methods Research  
Statistical Research Division

Subject: Clustering of Twenty-Eight Sample Areas in the  
Ethnographic Evaluation of the Behavioral Causes  
of Census Undercount for the 1990 Decennial Census

Attached is the first of two reports on results of the study whose objective is to assess the relative values of explanatory variables on the prediction of census omission and erroneous enumeration based on the data collected for the Ethnographic Evaluation of the Behavioral Causes of Census Undercount for the 1990 Decennial Census. This report focuses on the result of an explanatory data analysis in which the twenty-eight sample areas in the continental United States were grouped based on the data from the Sample Edited Detail File.

## **I. INTRODUCTION**

Based on data from the Ethnographic Evaluation of the Behavioral Causes of Census Undercount for the 1990 Decennial Census, de la Puente (1993a) examined the effects of demographic variables on census omission through use of the logistic linear regression models. One of his key findings was that inclusion of the sample areas in a logistic model was essential in obtaining a good fit to the data. However, it was not clear from the paper how the effect of the sample areas was introduced into the logistic model and how large the effect was in relation to the effects of the remaining explanatory variables.

This paper explores a way to define the effect of the twenty-eight sample areas located in the continental U.S. The approach taken is to group the sample areas into clusters, using the heretofore untapped site-level data on various aspects of the sample areas, including their social, economic, and educational backgrounds. The newly defined clusters will be included in later analyses as one of the possible factors that best explain the census coverage errors.

## **II. BACKGROUND**

The Census Bureau began a series of ethnographic evaluations of census coverage in 1986, culminating in the 1990 Ethnographic Evaluation. The history of the ethnographic evaluations and the study design of the 1990 Ethnographic Evaluation have been documented by Brownrigg and Martin (1989) and Martin, Brownrigg, and Fay (1990). The 1990 Ethnographic Evaluation consisted of intensive studies of twenty-nine small areas conducted by ethnographers. Each principal ethnographer had a close tie with the community and previously worked in and resided near the study area. As part of the evaluation project, each ethnographer conducted an Alternative Enumeration (AE) which was an independent (from the census) listing of the residents in the sample area, using participant observation, direct observation, and ethnographic interviews. The ethnographer collected Census Day residency status of each individual during the period of June to August 1990. The AE person list was later linked to the census person list at the Census Bureau, and persons missed or erroneously enumerated by the census were identified by the ethnographers in the Resolved Enumeration (RE). Each sample area included about 100 households in one or more census blocks. Twenty-eight of the sample areas were located in the continental U.S. and one in Puerto Rico. This paper will be concerned only with the twenty-eight sample areas in the continental U.S. The sample areas were selected, purposively, representing five groups

(Blacks, Hispanics, Asians, American Indians, and recent immigrants) in which undercounts were known or suspected to be high. The sample areas were also selected from three settings: ethnically homogeneous urban areas, ethnically heterogeneous urban and suburban areas, and ethnically homogeneous rural areas. In all, there were a total of 110 census blocks, 3367 housing units and 8718 individuals in the RE list. Table 1 in the attachment lists the sample areas and their locations. Figure 1 in the attachment shows the implemented sample design of the Alternative Enumeration by race/ethnicity and type of setting.

At the end of the ethnographic project, each ethnographer summarized his/her findings in an ethnographic coverage report (see de la Puente, 1993b). In addition, the ethnographer also filled out a "behavioral log" for each housing unit (HU) and selected individuals within his/her sample area. The behavioral log contained information on household structure, immigrant status and other housing unit and person-level information believed to be relevant to census coverage errors. The ethnographer was also asked to collect site-level data, pertaining to the entire neighborhood such as the presence of gangs and violence, employment rate, residential mobility. These data from the behavioral logs and ethnographers' observations were invaluable sources of information, not collected in the census, on the factors that the ethnographers mentioned in their coverage reports as important components of the multiple causes of the census coverage errors. However, the data were found to be lacking in completeness and consistency. Hence, these data were not used in the current or later data analyses.

Another source of the site-level data was available from the Sample Edited Detail File (SEDF) based on the census long-form questionnaires. The census long-form questionnaires were sent to one out of every six housing units, on the average, throughout the U.S. For the 1990 Ethnographic Evaluation, the SEDF data were available for the sample areas and one ring of blocks surrounding the sample areas (see Reference [9]). The data included aggregate information on social, economic, and educational backgrounds of the persons residing in and around the sample areas.

This paper attempts to answer following two questions: how can a sample area effect be summarized and defined? And how can the untapped SEDF data be utilized in the analyses of census coverage errors? The questions are answered jointly through grouping of the twenty-eight sample areas into clusters, based on the site-level data from the census long-form questionnaires in lieu of the behavioral logs and ethnographers' site-level observations. The result of the clustering will be examined against the

ethnographers' observations documented in the Ethnographic Coverage Reports.

Here, we are assuming that the data from the SEDF are representative of the sample areas investigated by the ethnographers in the Ethnographic Evaluation. However, this assumption might not be valid for some of the sites. For example, some site might have had a lot of persons missed by the census but enumerated in the AE. The discrepancy between the two person counts might be big enough to invalidate the social, economic, and educational environment of the site projected by the data from the SEDF. Another example is a sample area which might have included one or two large apartment buildings, containing an enclave of a particular ethnic group.

Hence, to check the representativeness of the SEDF data for the sample areas, the persons in the sample areas are compared with the persons who responded to the census long-form questionnaires in the ring blocks surrounding the sample areas through comparisons of some demographic variables for the two groups.

### **III. METHODOLOGY**

#### **A. DATA**

The data used in clustering the sample areas were the extracts of the Sample Edited Detail File. The SEDF contained information obtained from the responses to the census long-form questionnaires which were sent to a sample of all households in the U.S. Hence, the SEDF contained a probability sample of housing units and each person record in the SEDF had a person weight attached to it. Because there were substantial differences between the Puerto Rico and stateside detail files, the Puerto Rico site was excluded in this study. Also excluded from the study were 145 Group Quarters persons. Table 1 in the attachment shows the unweighted and weighted person counts from the SEDF for the sample areas and their ring blocks for each of the 28 sites in the study.

For comparison of the demographic variables, the Resolved Enumeration data were used for the sample areas in the Ethnographic Evaluation, and the SEDF was used for the ring blocks. Four sites (BEL, FRA, GAR and MOO) had no data for the ring blocks. Hence, these four sites were excluded from the comparisons.

#### **B. ANALYSIS METHOD**

The SAS's CLUSTER procedure was used to form hierarchical and disjoint clusters of the sample areas. Disjoint clusters place each sample area in one and only one cluster. The average linkage method (Sokal and Michener, 1958) was used to compute the distance between two clusters. All data were standardized. The pseudo F statistic, which measures the separation among all the clusters at each clustering level, was used to decide the number of clusters.

Clustering of the sample areas was based on the following eight variables derived from the SEDF data:

1. %LowEd = Weighted proportion of persons with less than high school education among persons 18 years old or older;
2. MedInc = Weighted median household income;
3. %Foreign = Weighted proportion of persons born abroad;
4. %OthLan = Weighted proportion of persons who spoke a language other than English at home;
5. %FemHH = Weighted proportion of female householders with no spouse;
6. %Owner = Weighted proportion of owner-occupied HUs;
7. %Vac = Unweighted proportion of vacant HUs\*; and
8. %Ereturn = Weighted proportion of households enumerated by enumerators, as opposed to by mail.

\* The proportion could not be weighted because the weights of vacant HUs were not known.

The %Ereturn variable was included in the study because the variable had been shown in earlier studies to be closely associated with the census undercoverage and overcoverage (see Griffin and Moriarity, 1992, and Moriarity and Childers, 1993).

The demographic variables that were used in comparing the sample areas in the Ethnographic Evaluation and their ring blocks from the SEDF were defined below:

RELATION: 1 = Householder  
 2 = Spouse  
 3 = Other relative of householder  
 4 = Not related to householder

GENDER: 1 = Male  
2 = Female

AGE: 1 = 17 years old or younger  
2 = 18-29 years old  
3 = 30-49 years old  
4 = 50 years old or older

MSTAT: 1 = Married  
2 = Never married  
3 = Separated/divorced/widowed

HISPANIC ORIGIN: 1 = Yes  
2 = No

#### **IV. LIMITATIONS**

Clustering of the sample areas was based on data with sparse observations in and around the sample blocks, and on a small number of variables.

As mentioned earlier, the clustering of the twenty-eight sites based on the SEDF assumed that the data were representative of the areas enumerated by the ethnographers.

#### **V. RESULTS**

##### **A. CLUSTERING OF TWENTY-EIGHT SAMPLE AREAS**

The result of cluster analysis indicated five clusters. Figure 2 in the attachment shows the result of hierarchical clustering as a tree structure. The diagram shows the order in which the sample areas were combined. Two clusters contained nine sample areas. One cluster contained five sample areas. Another cluster contained three sample areas. The last cluster contained two sample areas. Figure 1 in the attachment shows the five clusters in the context of the sample design of the 1990 Ethnographic Evaluation. Table 2 in the attachment lists the sample areas and their site-level variable values by cluster.

The characteristics of Cluster 1 included high proportions of persons born abroad (35%-79%) and high proportions of persons who spoke a language other than English at home (48%-84%). The high proportions for these two variables indicate the presence of a large number of recent immigrants in the sample areas. According to Figure 1, the sample areas in Cluster 1 included large populations of Hispanics and Asians. The proportion of persons

with less than high school education was, on the average, high in Cluster 1, with six of the nine sample areas having more than 50%. The sample areas in Cluster 1 had the median household incomes ranging from \$12,000 to \$17,000. The vacancy rate was less than 9% in every sample area in Cluster 1, resulting in the lowest cluster average among the five clusters. The proportion of owners was 18% or less in every sample area, except in the STE sample area (55%), in Cluster 1.

In Cluster 2, seven of the nine sample areas had 70%-90% black population. All nine sample areas in Cluster 2 had high proportions of female householders with no spouse (42%-78%) and high vacancy rates (11%-39%). Cluster 2 also had high proportions of persons with less than high school education, with four of the nine sample areas having more than 50%. The proportions of persons born abroad were less than 10%, except in the STR (32%) and WIN (35%) sample areas, in Cluster 2.

Four of the five sample areas in Cluster 3 came from rural sites, each of which was racially homogeneous. Cluster 3 as a whole, however, included various race and ethnic groups, containing two largely American Indian sample areas, two largely Hispanic sample areas and one racially heterogeneous sample area. The most prominent characteristic of Cluster 3 was the high proportions of owners (61%-92%) compared to other clusters, reflecting the rural settings of these sample areas. The proportions of female householders with no spouse were relatively low compared to the sample areas in other clusters, ranging from 14% to 40%.

Cluster 4 included three sample areas where we expected to enumerate a large proportions of Hispanic and Asian immigrants as in Cluster 1. In fact, the ethnographers reported in the Ethnographic Coverage Reports (de la Puente, 1993) presence of undocumented immigrants in large numbers in these sample areas. In the MAH and ROD sample areas, however, the proportions of persons born abroad were relatively low (23 and 24 percents), based on the census long-form questionnaires. The discrepancy between the SEDF and the Ethnographic Coverage Reports for the MAH and ROD sample areas might have resulted because of concealment of undocumented immigrants from the census. Whatever the cause of the discrepancy, here is an example where the data from the SEDF might not have represented some aspect of the two sample areas well.

One difference between Cluster 4 and the other four clusters was that the three sample areas in Cluster 4 were the only sample areas in the study with the median household income above \$25,000. In addition, Cluster 4 differed from Cluster 1, another cluster with a large number of Hispanic and Asian immigrants, in

that Cluster 4 had the lowest average proportion of persons with less than high school education while Cluster 1 had the highest average proportion among the five clusters.

The last cluster, Cluster 5, included two seemingly different sample areas. The most notable characteristic of Cluster 5 was that it contained two sample areas that were the only sample areas among the twenty-eight sample areas located in the List/Enumerate (L/E) areas. The L/E areas were sparsely populated areas where the U.S. Postal Service dropped off unaddressed census short-form questionnaires and enumerators picked them up. In the L/E areas, the enumerators always administered census long-form questionnaires at the time they picked up the short forms. Hence, %Ereturn was 100 percent in the two sample areas in Cluster 5. In the non-L/E areas where people mailed back their census questionnaires, %Ereturn corresponded to the proportion of census questionnaires that were sent to any of the census coverage improvement follow-up operations for non-response and reconciliation.

In both sample areas in Cluster 5, the proportion of persons born abroad was low at 1%. In the JOJ sample area, however, the proportion of persons who spoke a language other than English was 64%. Since the JOJ sample area contained many American Indians, the low value of FOREN and the high value of OTHLAN might have indicated widespread use of native American Indian tongue among the JOJ sample area residents. The proportions of persons with less than high school education were low (15%-27%) in both sample areas.

Descriptive cluster labels are given at the bottom of Figure 1 in the attachment.

Although none of the eight site-level variables used in clustering the sample areas was race/ethnicity or geography, the two factors of the Ethnographic Evaluation sample design, the cluster definition was closely related with the sample design. For example, Clusters 1 and 4 were confined to the Asian and Hispanic groups in urban and suburban areas. Clusters 3 and 5 were confined to rural areas.

And yet, the clusters were not defined strictly along the race/ethnicity and geography boundaries. For example, Cluster 2 included six of the eight sample areas in the Black category and the Hartford, CT (AMM) sample area in the Hispanic category, all of which included more than 70 percent Black based on the RE data. The South Saint Louis, MO (RYN) and the Chicago, IL (STR) sample areas had less than 10 percent Black population. However, these two sample areas had large proportions of female



householders with no spouse and of vacant units, just as the other seven sample areas in Cluster 2 did. Another example included Clusters 1 and 4, both of which had high concentrations of Hispanic and Asian immigrants. The sample areas in Cluster 1 had persons with less education than persons in Cluster 4, and their average median household income was less than that of the sample areas in Cluster 4.

Hence, use of the eight site-level variables in clustering appeared to fine-tune the cluster definition that would have been made based solely on the race/ethnicity and geography factors.

B. COMPARISON OF DEMOGRAPHIC VARIABLES BETWEEN SAMPLE BLOCKS IN THE ETHNOGRAPHIC EVALUATION AND RING BLOCKS IN THE SAMPLE EDITED DETAIL FILE

The overall frequency distribution of each of the five demographic variables for the sample blocks from the Resolved Enumeration was similar to the overall frequency distribution of the corresponding demographic variable for the ring blocks from the SEDF. The biggest difference was observed with respect to the Hispanic Origin variable. While 27.7 percent of the persons in the sample blocks were of Hispanic Origin, 32.7 percent of the persons in the ring blocks were of Hispanic Origin. The persons in the sample blocks had a slightly higher proportion of males (50.5% versus 47.8%) and they were slightly younger than the persons in the ring blocks.

When the frequency distributions were compared at each site, larger differences could be observed.

For each of the three Asian sample areas which were racially homogeneous urban sites (KAN, KIM, SUN), the ring blocks had a higher proportion of persons not related to the householder, a higher proportion of single persons and a higher proportion of persons of Hispanic Origin than the sample blocks (see Table 3). According to the Ethnographic Coverage Reports, these three sites consisted of apartment buildings with high concentrations of Asian residents who were elderly. This observation and the observed proportions indicate that each of the sites was an enclave of a particular Asian group embedded in the area that had a sizable population of Hispanics.

The STE and WIN sample areas included a large proportion of Haitian immigrants, according to the Ethnographic Coverage Reports. The comparison of the Hispanic Origin variable showed that the ring blocks had higher concentrations of Hispanics (29.3% in STE and 12.8% in WIN) than the sample blocks (3.5% in STE and 0% in WIN). Hence, each of these sites was apparently an

enclave of Haitian immigrants. The sample blocks for both the STE and WIN sites had higher proportions of single persons (15.5% more in STE, 18.0% more in WIN) than the ring blocks.

Several sample areas (BRA, MAH, MON, ROD, ROM) had much higher proportions of Hispanic population in the sample blocks than the ring blocks. These sample areas also had higher proportions of persons not related to the householder, single people and persons aged 19-39. The most extreme case was the ROD sample area (Houston, TX) where the sample block had 7.6 percent more 'non-relatives,' 16.6 percent more males, 18.3 percent more persons aged 19-39, 9.9 percent more single persons and 68.9 percent more Hispanics than the ring blocks (see Table 4). For these sample areas, the presence of undocumented immigrants observed by the ethnographers but apparently not enumerated by the census might explain the discrepancies, especially the discrepancy in the proportion of Hispanic population, between the sample and ring blocks.

Among the Black sample areas, the HAM sample area stood out, having large discrepancies in the demographic characteristics between the sample and ring blocks. The sample block had 16.2 percent more persons not related to the householders, 7.2 percent more males, 8.5 percent more persons aged 19-39, and 31.4 percent more single persons than the ring blocks. Also, the sample block had zero percent persons separated/divorced/widowed while the ring blocks had 23.3 percent. Hamid (1992) reported that this sample area, a section of one census block in central Harlem in northern Manhattan, NY, was scattered with deteriorated and abandoned buildings. He also observed that drug trafficking, drug use, and other crimes were chronic problems in the sample area as well as in the neighborhood that surrounded it. Hence, the discrepancy in the SEDF and the RE data here might have resulted from failure of the census and success of the ethnographic approach in enumerating individuals in the hard-to-enumerate areas.

In the JOJ sample area (Isleta Pueblo, NM), 37 percent of the residents were aged 18 and under, and yet, only 1.1 percent of the residents were single, based on the Resolved Enumeration. These inconsistent numbers cast some doubt on the quality of the AE data. In the JOJ ring blocks, 36 percent of the residents were aged 18 and under and 56 percent were single, based on the SEDF data.

Hence, the comparisons of the demographic variables showed that there were differences between the sample areas and their ring blocks. Some of the sample areas, for example, seemed to contain pockets of certain isolated ethnic groups. In the HAM sample

area, differences in the distributions of the demographic variables were observed which might have been caused by the differences in coverage between the census and the ethnographic procedure. However, it is not clear whether similar differences would be observed with respect to the eight site-level variables defined in Section III.B between the sample areas and the ring blocks. One would expect, in most cases, that the socioeconomic landscape would not drastically vary over an area covering several contiguous census blocks at a fixed point in time.

## **VI. CONCLUSION**

The eight site-level variables, derived from the census long-form questionnaires, were used in grouping the 28 sample areas into five clusters. The eight variables were crude measures of social, economic, and educational indicators of the sample areas. Although none of the variables were on geography or race/ethnicity, the results of clustering showed a remarkable alignment with the sample design of the 1990 Ethnographic Evaluation. At the same time, the extent that the cluster definition disagreed with the sample design implied the presence of additional information contained in the eight variables over race/ethnicity and geography.

The comparisons of the five demographic variables between the sample areas and the ring blocks showed that a fair amount of differences were observed between the population in the sample area enumerated in the AE and the population in the ring blocks enumerated by the census long-form questionnaires at the site level. However, whether similar differences might be observed with respect to the eight variables from the SEDF on the social, economic and educational backgrounds of the sample areas was not clear from this study.

The clusters formed in this study will be further examined in a later study for their effectiveness in predicting census omissions and erroneous enumerations.

## **REFERENCE**

1. Brownrigg, L. and Martin, E. (1989). "Proposed Study Plan for Ethnographic Evaluation of Behavioral Causes of Undercount." Paper presented to the Census Advisory Committee of the American Statistical Association and on Population Statistics at the Joint Advisory Committee Meeting, Alexandria, VA, April 13, 1989.
2. de la Puente, M. (1993a). "A Multivariate Analysis of the Census Omission of Hispanics and Non-Hispanic Whites,

- Blacks, Asians and American Indians: Evidence from Small Area Ethnographic Studies." Proceedings of the Survey Research Methods Section, American Statistical Association, 641-646.
3. de la Puente, M. (1993b). "Why Are People Missed or Erroneously Included by the Census: A Summary of Findings from Ethnographic Coverage Reports." Paper prepared for the Advisory Committee for the Design of the Year 2000 Census Meeting, Suitland, MD, March 3, 1993.
  4. Griffin, D. and Moriarity, C. (1992). "Characteristics of Census Error." Proceedings of the Survey Research Methods Section, American Statistical Association, 512-517.
  5. Hamid, A. (1992). Ethnographic Follow-Up of a Predominately African American Population in a Sample Area in Central Harlem, New York City: Behavioral Causes of the Undercount of the 1990 Census. Ethnographic Evaluation of the 1990 Decennial Census Report #11. Prepared under Joint Statistical Agreement 89-28 with John Jay College of Criminal Justice in New York City. Washington, DC: Bureau of the Census.
  6. Martin, E., Brownrigg, L. and Fay, R. (1990). "Results of 1988 Ethnographic Studies of Census Coverage and Plans for 1990." Paper presented to the Census Advisory Committees of the American Statistical Association and on Population Statistics at the Joint Advisory Committee Meeting, Alexandria, VA, October 18-19, 1990.
  7. Moriarity, C. and Childers, D. (1993). "Analysis of Census Omissions: Preliminary Results." Proceedings of the Survey Research Methods Section, American Statistical Association, 629-634.
  8. Sokal, R. and Michener, C. (1958). "A Statistical Method for Evaluating Systematic Relationships." University of Kansas Science Bulletin, 38, 1409-1438.
  9. U.S. Bureau of the Census. (1992). Memorandum for E. Martin from T. Walsh, dated May 27, 1992. "Delivery of Sample Edited Detail File Extracts for 1990 Ethnographic Evaluation Sites Requested as Part of the Behavioral Research Project." Washington, DC: Bureau of the Census.

TABLE 1.

**SAMPLE AREA CODE, ITS LOCATION,  
AND PERSON COUNT (UNWEIGHTED AND WEIGHTED)  
FOR THE SAMPLE AREA AND ITS RING BLOCKS:  
FROM THE SAMPLE EDITED DETAIL FILE**

Sample Area		P e r s o n C o u n t	
Code	Location	Unweighted	Weighted
AMM	Hartford, CT	282	3176
ASH	N St Louis, MO	85	887
BEL*	Logan County, OK	214	556
BRA	New Orleans, LA	240	2008
BRI	Carbondale, IL	78	586
BUN	Long Beach, CA	322	3156
DAR	Flint, MI	137	886
DOM	Bronx, NY	943	8716
DUR	Orleans Parish, LA	121	946
FRA*	Holmes County, MS	445	1354
GAR*	Santa Barbara, CA	318	1804
HAM	Harlem, NY	496	5255
JOJ	Isleta Pueblo, NM	473	1014
KAN	Queens, NY	990	9540
KIM	Koreatown, CA	602	5032
LER	Little Branch, NC	294	1500
MAH	Long Island, NY	46	336
MON	Marion County, OR	139	1279
MOO*	Okfuskee County, OK	186	1016
ROD	Houston, TX	821	8418
ROM	San Francisco, CA	305	2964
RYN	S St Louis, MO	261	2394
SHA	North Beach, CA	124	984
STE	Miami, FL	188	1429
STR	Chicago, IL	280	3048
SUN	Chinatown, NY	1307	10781
VEL	San Diego, CA	219	1492
WIN	Ft Lauderdale, FL	90	838
Total		10006	81395

\*Note: Only sample block data (and, therefore, no ring block data) were available for the BEL, FRA, GAR and MOO sample areas.

FIGURE 1.

**FIVE CLUSTERS OF TWENTY-EIGHT SAMPLE AREAS  
IN THE FRAMEWORK OF  
THE 1990 ETHNOGRAPHIC EVALUATION SAMPLE DESIGN,  
BY RACE/ETHNICITY AND TYPE OF SETTING**

(Three-letter abbreviation is the sample area code.)

Race/ Ethnicity	Racially Homogeneous Urban Site		Racially Heterogeneous Urban/Suburban		Racially Homogeneous Rural Site	
	clstr no.	sample area	clstr no.	sample area	clstr no.	sample area
ASIAN	1	SUN	1	BUN		
	1	KIM	1	SHA		
	4	KAN	2	RYN		
			2	STR		
Recent Immigrants	1	DOM	1	ROM		
	1	STE	4	MAH		
			4	ROD		
HISPANIC	1	VEL	1	BRA	3	GAR
			2	AMM	3	MON
BLACK	2	DAR	2	ASH	2	FRA
	2	DUR	3	BRI	5	BEL
	2	HAM				
	2	WIN				
AMERICAN INDIAN					3	LER
					3	MOO
					5	JOJ

Note: 1. The Puerto Rico sample area (DUA) is not included in this table.  
2. Refer to Table 1 for the relation of the sample area code to its location.

Cluster Labels (A number in parentheses indicates the number of sample areas.):  
Cluster 1: Hispanic and Asian immigrants with low household income (9);  
Cluster 2: Blacks (9);  
Cluster 3: American Indian/Hispanic rural homeowners (5);  
Cluster 4: Hispanic and Asian immigrants with high household income (3);  
Cluster 5: List/Enumerate sample areas (2).

TABLE 3.

**COMPARISON OF DEMOGRAPHIC VARIABLES  
BETWEEN THE SAMPLE AREA AND ITS RING BLOCKS  
FOR KIM (KOREATOWN, CA) SAMPLE AREA**

	Percent Distribution	
	Sample Block	Ring Blocks
<hr/>		
RELATION		
Householder	43.7	27.5
Spouse	23.2	11.0
Other Relative	33.1	47.0
Non-Relative	0.0	14.5
Total	100.0	100.0
AGE		
18 or under	17.2	29.1
19-39	21.2	44.6
40 or above	61.6	26.3
Total	100.0	100.0
GENDER		
Male	44.4	51.9
Female	55.6	48.1
Total	100.0	100.0
MSTAT		
Married	53.0	28.6
Never Married	29.1	61.8
Div/Sep/Wid	17.9	9.6
Total	100.0	100.0
HISPANIC ORIGIN		
Yes	6.0	81.8
No	94.0	18.2
Total	100.0	100.0

Source: Resolved Enumeration Data for the sample area and  
Sample Edited Detail File for the ring blocks

TABLE 4.

COMPARISON OF DEMOGRAPHIC VARIABLES  
BETWEEN THE SAMPLE AREA AND ITS RING BLOCKS  
FOR ROD (HOUSTON, TX) SAMPLE AREA

	Percent Distribution	
	Sample Block	Ring Blocks
-----		
RELATION		
Householder	35.0	61.0
Spouse	15.5	15.2
Other Relative	35.8	17.7
Non-Relative	13.7	6.1
Total	100.0	100.0
AGE		
18 or under	19.0	9.7
19-39	72.6	54.3
40 or above	8.4	36.0
Total	100.0	100.0
GENDER		
Male	66.4	49.8
Female	33.6	50.2
Total	100.0	100.0
MSTAT		
Married	41.6	33.4
Never Married	55.7	45.8
Div/Sep/Wid	2.7	20.8
Total	100.0	100.0
HISPANIC ORIGIN		
Yes	86.3	17.4
No	13.7	82.6
Total	100.0	100.0

Source: Resolved Enumeration Data for the sample area and  
Sample Edited Detail File for the ring blocks