UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

This memorandum is intended for internal Census Bureau use and for external stakeholders who are interested in the Census Bureau's continuing research efforts. If you have any questions regarding the use or dissemination of this information, please contact Tasha Boone, Assistant Division Chief for Census 2010 Publicity Office (C2PO), ADCOM Directorate, at (301) 763-3977.

October 24, 2008

C2PO 2010 Census Integrated Communications Research Memoranda Series

No. 1

MEMORANDUM FOR	Distribution List
From:	Tasha Boone *[signed]* Acting Chief, Census 2010 Publicity Office
Subject:	Segmenting the Population for the Census 2010 Integrated Communication Program
Contact Person:	Nancy Bates, C2PO Lead Researcher, (301) 763-5248

Attached is the C2PO 2010 Census Integrated Communications Research of Segmenting the Population for the Census 2010. This research defines the underlying constructs behind the hard-to-count mailback populations, develops mutually exclusive clusters of the population according to mailback propensity, and models the potential impact that the partnership and advertising campaign may have on mail response among these population clusters.

Attachment

Segmenting the Population for the Census 2010

Integrated Communications Program[1]

Nancy Bates, Census 2010 Publicity Office

Mary H. Mulry, Statistical Research Division

October 22, 2007

Introduction

During Census 2000, the Census Bureau hired the advertising firm of Young and Rubicam, Inc. (Y&R) to develop and deliver persuasive advertising and messaging designed to increase census awareness and mail response. In order to segment the market, Y&R created a behavioral Likelihood Spectrum™ Model designed to predict census participation. The central organizing assumption behind the model was that participation in civic and community-minded activities could also predict participation in the census. The model collapsed the population into three groups using civic-mindedness as a proxy to predict census participation. These included the least likely to respond (17% of the population), undecided/passive (43% of the population) and the most likely to respond (40% of population) (see Baron and Billia, 1999).

As part of the 2010 Census communications contract, the vendor is again expected to develop a campaign that provides a research-based strategy including "an audience segmentation framework to be used as the basis for creative direction and media strategy" (U.S. Census Bureau, 2006). In order to assist the contractor, the Census Bureau has

[1] Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are the authors' and not necessarily those of the U.S. Census Bureau. Do not quote or cite without permission.

begun a series of research projects designed to segment the population in ways that will help the vendor target their markets and design and deliver media strategies. In this paper we present findings from a macro-level segmentation study. First, we define the underlying constructs behind the hard-to-count mailback populations, second, we develop mutually exclusive clusters of the population according to mailback propensity, and third, we model the potential impact that the partnership and advertising campaign may have on mail response among these population clusters.

Segmenting the market: macro vs. micro

Similar to the Likelihood Spectrum™ used in 2000, we propose that the population is made up of three broad segments: those likely to respond by mail, those undecided, and those unlikely to respond by mail. To identify these segments, a variety of data on mail response behavior and survey participation in general is available for analysis. One excellent data source available at the macro-level is the Tract-Level Planning Database with 2000 Data (PDB). The PDB is a database[2] that assembles a range of housing, demographic, and socioeconomic variables correlated with mail nonresponse (Bruce and Robinson, 2007). These data allow for segmentation of the population according to indicators related to mailback behavior.

In addition to the PDB analysis, micro-level data sources are also available to validate and supplement the macro-level model. For example, the Census 2000 100% Detail File,

[2] The PDB includes all tracts with population and housing units in the Census 2000 mail universe (mailout/mailback and list enumerate).

100% Edited Detail File, and the American Community Survey (ACS) files contain household-level characteristics related to mail return behavior. Analysis of the 100% files will yield information about the types of households that mailed back Census forms immediately versus those that waited until weeks later when the motivational advertising phase commenced. Likewise, an examination of household characteristics by mode of response in the ACS will profile the type of households that respond early (i.e., completes the original ACS form) versus those that respond later to the replacement form versus those who do not respond by mail at all. In addition, a special ACS study conducted in 2004 captured the "person number" of the individual within the household who filled out the ACS questionnaire. This allows us to isolate the person-level characteristics such as age, relationship, and sex of the "form fillers" among different types of households (e.g., spousal households, non-related households, linguistically isolated households, etc.).

Additionally, data from Census 2000 exist regarding the relationship between mail response and attitudinal/behavioral measures like privacy and confidentiality concerns, trust in government, and civic and political participation. These data can be found in the 2000 Census Monitoring Surveys conducted by InterSurvey and the 2000 Partnership and Marketing Program (PMP) surveys conducted by the National Opinion Research Center. Finally, paradata from the Contact History Instrument (CHI) used in some personal-visit Census Bureau demographic surveys can provide insight into the reasons why some households participate in surveys while others do not (Bates, Dahlhamer and Singer, 2006; Bates and Piani, 2005). This broad approach of tapping both macro and micro-level data attempts to answer both the "who" and the "why" behind response behavior.

In this paper we present results from a macro-level analysis of the 2000 PDB – the micro-level research will be issued in separate reports.

Macro Level Analysis: The 2000 Planning Data Base (PDB)

Factor Analysis

The ultimate goal of the mailback population segmentation is to convey meaningful information to the contractor so they may, for example:

- Validate and supplement the segmentation model with geographic marketing databases (e.g., Claritas, PRIZM)
- Validate the attitude and opinion models with consumer psychographic databases (e.g. Survey of American Consumer, Simmons Research)
- Develop and test messages using attitudinal and behavioral data.
- Target and place media according to each of the underlying factors and clusters identified.
- Make resource allocation decisions (e.g., where to put Partnership vs. advertising and what populations *not* to target or expend resources because mail response is already high).

In addition to housing and socioeconomic indicators, the PDB also contains a Hard-to-Count (HTC) score. This score is highly correlated with mail return rates and is constructed from twelve variables:

- % vacant units,
- % non-single family attached/detached units;
- % renter occupied units;
- % units with >1.5 persons per room ;
- % non-spousal units;
- % units without phone;
- % people below poverty level;
- % units receiving public assistance;

- % people unemployed;
- % linguistically isolated households,
- % moved within last year, and
- % without a high school degree.

We can use the HTC scores to simply see *where* the tough tracts are, that is, the tracts with low mail return rates relative to the national average in 2000. We can also easily identify which variables correlate most strongly with mail return rates and contribute most to the HTC scores. But, what are the underlying constructs behinds the tracts with below average mail return rates? Not all tracts with low mail response have the same sociodemographic characteristics. Consequently, we need a more focused segmentation to aid the contractor and partnership program as they develop messages and target activities, media, and advertising.

To address this research question, we first conducted a factor analysis to deconstruct and identify a smaller number of unique factors underlying the 12 variables that compose the HTC score.[3] The resulting factors represent distinct and non-correlated "snapshots" of the hard-to-count. We performed an exploratory principal components factor analysis (with varimax rotation) using the 2000 PDB[4]. Once the factors were extracted, we next computed scores for each tract using the NFACT option with the SAS FACTOR procedure. These scores determine the optimal regression weights, multiply the tract

[3] We used only the 12 variables used to compute the HTC score. The additional PDB variables are highly correlated with these 12 and do not appear to add much information. For this analysis we also merged a tract-level mail return rate variable with the PDB. These rates are not identical to the "official" mail return rates reported from the 2000 Census.

[4] The 2476 tracts flagged as "nonrepresentative" were excluded from the factor analysis. These represent about 3.7% of the 2000 PDB tracts. These are sparsely populated tracts containing populations of less than 250 and/or less than 100 housing units; tracts with group quarters population greater than 50%, and tracts containing 35% or greater population aged 65+.

characteristics by the weights, and sum the products. This assigns a score for each tract to indicate how strongly it is related to each construct.

The analysis revealed three distinct factors that set the foundation for understanding populations with low mail return in 2000 (see appendix A for factor loadings). We subsequently labeled these factors:

(1) The Economically Disadvantaged,

(2) The Unattached/Mobile Singles, and

(3) High Density Areas with Ethnic Enclaves.

The Economically Disadvantaged factor had high loadings on vacant housing, poverty, public assistance, unemployment, less than a high school education, and absence of a phone (see Table 1 for a summary). This factor explains the largest portion of the variance and has the largest negative correlation with mail return rates (-.56). The average mail return rate in tracts scoring high on this factor was far below average at 63.5 percent (the overall average tract mail return rate was 75.4%). Tracts scoring high[5] on this factor also had a high correlation with % Black and a moderate correlation with % American Indian and Alaskan Native (AIAN). Tracts with high Economically Disadvantaged scores had an average HTC score of 75 (well above the overall average tract HTC score of 33). In summary, this factor reflects a struggling underclass population and underserved communities.

[5] Tracts with a score of 1.25 or higher for Factor 1 were defined as "high" for Factor 1 (11% of all tracts); tracts scoring 1.75 or higher on Factor 2 were defined as "high" for Factor 2 (7% of all tracts); and tracts scoring 1.75 or higher on Factor 3 score were defined as "high" for Factor 3 (6%).

The second factor (Unattached/Mobile Singles) is distinct from the first with high factor loadings on non-spousal households, renters, multi-unit structures, and residential mobility within the last year. Tracts loading high on the Unattached/Mobile Single factor also had below average mail return rates (66.5%) and a fairly strong negative correction with mail return rate (-.48). Tracts closely aligned with this factor did not indicate a strong correlation with any one race or ethnic group. In summary, this factor tends to reflect mobile, single adults, many of whom do not have children and may be living on their own for the first time.

The final factor (High Density w/Ethnic Enclaves) loaded high on only three HTC variables: crowded housing, linguistic isolation, and less than high school education. Tracts with high scores on this factor had below average mail return rate (67.2%), an above average HTC score (75), a strong correlation with % Hispanic and some correlation with % Asian or Native Hawaiian/Other Pacific Islander (NHOPI). The underlying construct with this factor appears to be densely populated ethnic enclaves -- some with limited English language proficiency. It is noteworthy that the three factors that emerged from our study are very similar to those documented by Bruce (2003) in a county-level factor analysis.

In summary, our factor analysis groups *variables* into distinct underlying factors – in our case we use the twelve variables that make up the HTC score. The analysis revealed three noncorrelated dimensions (sometimes referred to as unobserved variables) highlighting three different population segments all hard to count by mail. This serves as the

foundation for understanding the below average mailback population and how they represent three distinct constructs.

Macro Level Analysis: Cluster Analysis

Following the factor analysis, we performed a cluster analysis also using data from the 2000 tract-level Planning Data Base.[6] Unlike factor analysis, a cluster analysis groups *objects* (in our case tracts) with similar characteristics into relatively homogenous subsets. The cluster analysis groups each and every tract into one of several mutually exclusive clusters creating a multidimensional classification typology. The goal is to produce a macro-level market segmentation based on propensity to mail back a Census 2000 form. Unlike the factor analysis which serves to illustrate the underpinnings of the hard-to-count populations, the cluster analysis encompasses the entire spectrum of mailback propensities from high mail return rates to low. The two techniques are complimentary since both perform clustering functions, but with slightly different purposes.

There are many ways to perform cluster analysis. Our study uses the SAS procedure FASTCLUS to perform a disjoint cluster analysis based on distances computed using the 12 Hard-to-Count score variables in the PDB. Each observation (i.e., a tract) is assigned to one and only one cluster. The FASTCLUS procedure uses Euclidean distances so the cluster centers are based on least-squares estimation. The method is sometimes called the

[6] The data source for the cluster analysis was the PDB merged with an extract from the Census 2000 Summary File 1 and geography records. The latter extract provided additional tract-level measures of urbanicity and population density.

k-means model, since the cluster centers are the means of the observation assigned to each cluster.

For our analysis, we requested eight mutually exclusive clusters and a maximum number of 100 iterations.[7] We settled upon these parameters after several rounds of exploratory analysis using fewer clusters and iterations. Eight clusters seemed to satisfy our requirements by producing distinct enough groups that could be logically named according to their differences from (and in some cases similarity to) one another. The eight groups ranged in size from the largest (representing 35% of all occupied housing units) to the smallest (reflecting only 2% of all occupied units).

Below is a description of the 8 clusters (see Appendix B for the unweighted cluster means[8] and Table 2 for a summary):

Cluster 1: *All Around Average I (homeowner skewed)*

Mail Return Rate: 77%
Hard-to-Count score: 23
36.5 million occupied HUs (35% of total)
21,174 tracts (34% of total)

This group had the second highest mail return rate in 2000. They are best described as the "average Joe" cluster in that they are close to average on every one of the HTC variables. Around 28% of the housing units are not single-family structures, only one-quarter are renters, and slightly less than half (45%) are in non-spousal households.

Unemployment, poverty, education and mobility levels are all close to national averages. The tracts are fairly representative of the national average racial breakouts but have above-average percentage of non-Hispanic Whites (80%) slightly below-average Blacks

[7] The algorithm converged in 9 iterations.
[8] We also calculated the weighted means based on the number of occupied households per cluster -- the difference between weighted and unweighted was not large enough to change any inferences.

(9%), 2% Asian or NHPI and 1% AIAN. Tracts in this cluster contain about 7% Hispanics which is below the national average. Around one-quarter of the population is under age 18 and about 15% are over 65.

This group is the largest cluster representing about 35.5 million occupied housing units (about 35% of the total). This cluster has the largest percentage of rural tracts[9] (on average around 37% are rural). Tracts in this cluster may not require much targeting beyond the broad campaign elements designed to hit all sectors of the population. On the other hand, since this cluster includes so much of the population, even small increases in response rate will yield a very large number of mail returns.

Cluster 2 - *All Around Average II (renter skewed)*

Mail Return Rate: 74%
Hard-to-Count score: 41
16.5 million occupied housing units (16% of total)
8957 tracts (15% of total)

This cluster is also somewhat unremarkable and "average" on most of the HTC variables. About the only distinguishing characteristic is an above average number of households renting and in multi-units. This group of tracts is slightly more racially diverse than Cluster 1 (12% Black, 11% Hispanic, and 69% non-Hispanic White) and is also much more urban and densely populated.

Like Cluster 1, it may be wise to limit the amount of targeted resources invested into this cluster since MRRs are already average. However, like Cluster 1, this group is relatively large (represents around 16% of all occupied housing units). Taken together, Cluster 1 and 2 represent just over half of all occupied housing units. Consequently, even modest percentage gains in response will translate into a large absolute number of mail returns (and thus cost savings from large reductions in personal visit follow-ups).

Cluster 3 – *Economically Disadvantaged I (homeowner skewed)*

Mail Return Rate: 66%
Hard to Count score: 65
6.6 million occupied HUs (6% of total)
5,230 tracts (8% of total)

This cluster reflects households that are economically disadvantaged, but not as much as Cluster 4. One noticeable difference is that this cluster has fewer renters than Cluster 4

[9] "Urban" is defined as housing units located within urbanized areas (UAs) or urban clusters (UCs). A UA consists of areas containing 50,000 or more people while a UC consists of areas with a least 2,500 people but fewer than 50,000. "Rural" consists of areas located outside of UAs and UCs (U.S. Census Bureau, 2001).

(less than half rent – 46%). Nonetheless, these tracts have a high percentage in poverty, without a high school education, and on public assistance. Above average unemployment is also characteristic of this cluster.

Blacks comprise about one-half (49%) of the population in these tracts – the second largest Black population next to Cluster 4. This cluster has above-average number of children (29% are younger than 18).

This group represents about 6% of the total occupied housing units. The overwhelming majority of tracts in this cluster are urban (92% urban on average). This cluster will likely require targeting and special attention to achieve the desired mailback cooperation. Tracts in this cluster are probably good candidates to target activities of the partnership campaign (e.g. promoting Census participation through trusted community leaders).

Cluster 4 – *Economically disadvantaged II (renter skewed)*

Mail Return Rate: 58%
Hard-To-Count Score: 92
3 million occupied housing units (3% of total)
2574 tracts (4% of total)

This cluster had the lowest mail return rate of any group and also the highest HTC score. Close to three-quarters of the households in these tracts contain non-spousal renters in multi-units (especially 10+ units). These tracts also have the highest poverty, public assistance, and unemployment than any other cluster. This cluster most closely resembles Cluster 3 but has far fewer homeowners (on average, 81% of households are rented). Like Cluster 3, this group contains a higher percentage of Blacks (54%) but also has above-average percentage of Hispanics (21%).

This cluster reflects the most urban of all clusters (99.9% urban on average). This cluster represents about 3% of the total occupied housing units. Because this cluster contains above average percentage of minorities, a successful mailback campaign in these tracts could help decrease the differential undercount. Like Cluster 3, these tracts will likely require promotional activities that extend beyond paid advertising and other traditional media outlets.

Cluster 5 – *Ethnic enclave I (homeowner skewed)*

Mail Return Rate: 70%
Hard-To-Count Score: 63
3.4 million occupied HUs (3 % of total)
2440 tracts (4% of total)

This cluster is characterized by above-average crowding and poverty, public assistance, unemployment and low education. However it also contains a *below-average* percentage of non-spousal households and above-average percentage of children. It looks most like Cluster 6 with the following differences: lower occurrence of linguistic isolation, lower mobility, higher homeownership, and fewer Asians. This cluster is also less urban and less densely populated than Cluster 6. This group is predominantly Hispanic (61%) with 24% non-Hispanic White, 8% Black, and 5% Asian or NHOPI.

Like cluster 6, this cluster is also a candidate for in-language targeting and messaging. The make-up of this cluster indicates that gains in mail response rates could help to decrease the differential undercount. This group contains about 3% of all occupied housing units.

Cluster 6 – *Ethnic enclave II (renter skewed)*

Mail Return Rate: 64%
Hard-To-Count Score: 84
2.5 million occupied HUs (2% of total)
1754 tracts (3% of total)

This cluster has the second-highest HTC score of any cluster (and second lowest mailback rate). This cluster has above-average presence of children and is characterized by multi-unit structures with 10+ units. This group is exclusively urban, the most densely populated of clusters, and characterized by crowded housing. On average, half of persons residing within this cluster lack high school degrees. These tracts are predominantly comprised of Hispanics (59%) and Asians (11%) with only 19% non-Hispanic White, 9% Black and 1% AIAN.

This cluster contains tracts with high levels of linguistic isolation (on average, around 31%). In some tracts, this ranges as high as 79% of households where Spanish is spoken at home or no household member 14 or older speaks English very well. Likewise, other tracts have as high as 74% of households where an Asian/Pacific Islander language is spoken at home or no household member over 14 speaks English very well. This group is overwhelmingly renters (75%). It also has high poverty rates, unemployment, and public assistance.

This cluster is a candidate for in-language targeting and messaging and will likely require intense partnership efforts to gain trust and cooperation. This is the smallest of the 8

clusters representing only 2% of the total occupied housing units. As such, increases to response rates will yield a smaller number of actual mail forms compared to the other clusters. However, because this cluster contains a large number of Hispanics and Asians, increases to mail cooperation could translate into decreases in the differential undercount – one of the three stated goals of the Census 2010 communications campaign.

Cluster 7 - *Single/unattached/mobiles*

Mail Return Rate: 67%
Hard-to-Count score: 61
8 million occupied housing units (8% of total)
4,073 tracts (7% of total)

This cluster had a similar mail return rate and HTC score as Cluster 3 but looks very different. The overwhelming majority of households are non-spousal renters located in multi-units (especially structures with more than 10 units). These tracts have higher than average education, very high mobility, are densely populated, and almost exclusively urban. These tracts have below average percentage of children (17%). This cluster has a relatively high percent of group quarters (4%) – possibly reflecting college campuses. These tracts probably include younger singles in school or just out of school and into the workforce for the first time.

This cluster is racially diverse with above-average percent Asian (7%) and the majority non-Hispanic White (59%) followed by Black (17%). This group represents about 8% of the total occupied housing units. It may pay to target this cluster with media aimed for younger markets (e.g., digital media).

Cluster 8 – *Advantaged homeowners*

Mail Return Rate: 83%
Hard-to-Count score: 6
26.8 million occupied HUs (26% of total)
16,506 tracts (26% of total)

This group of tracts had the highest mail back rate and lowest HTC score in 2000. As such, these tracts have a very low percentage of renters, few multi-units structures, very low poverty levels and unemployment, low mobility and few non-spousal households. This cluster is indicative of stable homeowners who reside in spousal households in single unit houses, about one-quarter of which are located in non-urban areas. This group of tracts is the least racially diverse of all clusters with 85% non-Hispanic White and only 4% Black, 5% Hispanic, 4% Asian or NHOPI and less than 1% AIAN. It is also the least densely populated cluster as measured by population per square mile.

Tracts in this cluster appear historically primed to cooperate and may only require Census awareness and a few reminders to repeat the high levels documented in 2000. This group is the 2nd largest behind Cluster 1 reflecting 26% of the total occupied housing units.

Table 3 presents a slightly different perspective of the clusters by illustrating how race and ethnicity populations are distributed among them. Just over half of the non-Hispanic White population fall into tracts assigned to the All Around Average clusters I and II (54.5%) while just under half of the AIAN and NHOPI population falls into either of these two clusters (49.6% and 49.1%, respectively). Another 20% of the NHOPI population are located in the Advantaged Homeowner cluster. The remaining AIAN population was fairly evenly divided among the Economically Disadvantaged I, Ethnic Enclave I, and Advantaged Homeowner clusters (15.0%, 12.2% and 12.4%, respectively).

Around one-third of the non-Hispanic White population and just over one-quarter of the Asian populations reside in tracts assigned to the Advantaged Homeowner cluster. A notable percent (13%) of the Asian population were located in the Single/Unattached/Mobile cluster. Around 39% of the Black population resides in tracts assigned to All Around Average I or II while another 37% reside within tracts assigned to the Economically Disadvatanged I or II clusters. Around 37% of the Hispanic population was located in tracts assigned to the Ethnic Enclave I or II clusters with another one-third in either All Around Average I or II. Looking down the column of clusters, it appears that

All Around Average II and the Single/Unattached/Mobiles are the most racially and ethnically diverse clusters.

The cluster analysis revealed eight distinct groups each with varying levels of mail return behavior in 2000 and each with unique demographic, housing, and socioeconomic characteristics. In three instances, pairs of clusters appear closely related to one another with homeownership/renter status as the distinguishing feature (i.e., All Around Average I and II; Economically Disadvantaged I and II; and Ethnic Enclave I and II). Several of the clusters exhibit characteristics of the underlying factors uncovered in the earlier analysis (i.e., Economically Disadvantaged I and II; and Ethnic Enclave I and II; and Single/unattached/mobiles).

To illustrate the spatial separation of the clusters, Figure 1 depicts a canonical discriminant analysis on the clusters where the 12 variables are reduced to two canonical variables. These canonical variables are linear combinations of the 12 variables and are chosen to provide maximum separation between the clusters although the interpretation of the canonical variables is not easy. The canonical variables summarize the between-cluster variation similar to how principal component analysis summarizes total variation. For example, the Young/mobile/single cluster is closer to the All Around Average II - Renter Skewed cluster and the Economically Disadvantaged II - Renter Skewed cluster than to the rest of the clusters.

A second figure provides a geographic view of the eight clusters through a map of the U.S. with the tracts in each cluster shown in a different color (Figure 2). The All Around Average, the Economically Disadvantaged, and the Ethnic Enclave pairs of clusters each have a separate color with the renter-skewed cluster denoted by a darker shade. The Advantaged Homeowners and the Young/mobile/single clusters each have their own color. The map vividly illustrates how certain clusters tend to be concentrated in a particular geographic area. For example, the Advantaged Homeowners (blue) are particularly noticeable in the Midwest and along the North Atlantic seaboard while clusters of the Economically Disadvantaged I (purple) are apparent in the South especially along the Arkansas/Mississippi boarder. As expected, some of the Ethnic Enclave clusters (green) show up in California, Arizona, New Mexico and Texas.

Some of the clusters contain tracts that do not show up well on standard 8.5 by 11 inch paper since many of the tracts in these clusters are found in urban areas. However, zooming in on particular areas is possible with an electronic file of the map.

We expected that tracts with a high score on Factor 1 would tend to fall within the Economically Disadvantaged I or II clusters, that tracts with high scores on Factor 2 would fall within the Single/unattached/mobile cluster and that tracts scoring high on Factor 3 would fall into either of the two Ethnic Enclave clusters. To validate this, we performed a crosstabulation of the tracts with high factor scores by the eight clusters. Table 4 presents the results of these crosstabulations.

Two out of three tracts scoring high on the Economically Disadvantaged factor fell into either the Economically Disadvantaged I or II clusters (66%). While not a perfect one-to-one match, the inferences from the factor and cluster analysis generally support one another. Further validation that Factor 1 is capturing the construct of the underclass is evidenced by the fact that 0% of tracts scoring high on Factor 1 fell into the Advantaged Homeowner cluster. As expected, the majority of tracts in the Single/unattached/mobile cluster also had a high score on Factor 2 (71%). The remaining tracts scoring high on Factor 2 tended to fall into the Economically Disadvantaged II (renter skewed) cluster. Finally, the crosstabulations validate the notion that Factor 3 represents areas containing immigrants, particularly Hispanic and Asian immigrants. Approximately 81% of the tracts scoring high on this factor fell into either the Ethnic Enclave I or II clusters (39% and 42% respectively).

Cluster Segments and Increases in Mail Return Rates

The next step estimates the increase for each cluster in the number and percentage of the Census 2000 forms mailed back due to the influence of the 2000 Partnership and Marketing Program (PMP). The PMP consisted of the paid advertising and the partnerships with businesses, governments, and other organizations to promote the census. Young & Rubicam, Inc. (Y&R) planned and conducted the paid advertising campaign under contract with the Census Bureau.

The Y&R campaign included messages for an English-speaking audience with additional advertising focused toward populations who speak other languages and populations believed to be hard-to-count. The campaign had three phases:

- The educational phase began November 1, 1999 and lasted until January 30, 2000. The goal was to teach the public about the census.
- The motivational phase began February 28, 2000 and lasted until April 9, 2000. The bulk of the advertising was during this period. The goal was to encourage participation in the census.
- The nonresponse followup phase began on April 17, 2000 and lasted until mid-June. The ads informed the public that enumerators would be visiting to collect census data.

The methodology for estimating the increase in response due to PMP uses logistic regression models that estimate the probability of mailing back a census form (Mulry and Keller 2007). The models were fit using data collected in a survey to assess the public's awareness of the census in combination with a check in the census databases to determine if the respondents returned a census form by mail. The National Opinion Research Center (NORC) conducted the survey and logistic regression modeling as part of the Census 2000 Evaluation Program and described in a report by Wolter et. al. (2003).

First, the estimation of the increase in response due to the PMP uses the logistic regression model for the log-odds of a mail return to estimate the probability of returning a form for those who reported not having seen any of the communications. Then the

model is used to estimate the probability for those who have "average" scores on seeing the communications. Multiplying the difference in these probabilities by the population size produces an estimate of the number of additional returns.

NORC collected the data about the public's awareness of the census and the communication vehicles being used to promote the census by conducting three surveys, called waves, at three different times. The first wave was prior to most of the advertising and functioned as a baseline to estimate the increase in awareness of the census in the wave conducted prior to the forms being mailed and the wave conducted during Nonresponse Followup. The sample design included four different samples for each wave. The core sample covered the total population and allowed separate analyses of Hispanic, non-Hispanic Blacks, and Non-Hispanic White populations. In addition, a separate sample was selected for each of three other populations, Asians, Native Hawaiians, and American Indians and Alaska Natives (AIAN).

- The Wave 1 survey collected data from September 1, 1999 until November 13, 1999 and completed 3,002 interviews.
- The Wave 2 survey collected data from January 17, 2000 until March 11, 2000 and completed 2,716 interviews of which 1,193 were in a sample of the entire population, known as the core sample, and eligible for a mail return.
- The Wave 3 survey collected data from April 17, 2000 until June 17, 2000 and completed 4,247 interviews, 1944 of which were in a sample of the entire population, known as the core sample, and eligible for a mail return.

The respondents in Waves 2 and 3 were linked to the census databases to determine if a mail return was received for them. There is no model for Wave 1 because the link between the survey data and the census databases was attempted but proved problematic. An address was sent to Nonresponse Followup if the Census Bureau had not received a mail return by April 18. The definition of a mail return used by NORC attempted not to confound the measurement of the influence of PMP with the additional influence of being contacted by a Nonresponse Followup interviewer. For Wave 2, a form "was classified as a mail return if it had a valid census mail return date that was prior to the Nonresponse Followup interview date (NRD) provided on the Census Bureau file." (Wolter et al 2007, p.87) For Wave 3, "a mail return must have occurred before the NORC interview date and the NRD." (Wolter et al 2007, p. 87) The implication is that some of the Wave 2 or Wave 3 respondents designated as mailing back a form may have been sent for a Nonresponse Followup interview. The awareness measured for the Wave 3 respondents may have been influenced by advertising that occurred early in Nonresponse Followup. Contacting the Wave 2 respondents prior to their receipt of the census questionnaire may have conditioned them to being more likely to mail back their form.

Table 5 shows the estimated increase in the mail return rate and numbers due to PMP for the eight clusters. These estimates were derived by applying the Wave 2 models and using the average awareness of PMP measured within subgroups in Wave 3. The models included variables for race and Hispanic ethnicity. The estimation for Hispanics, Non-Hispanic Blacks, and Non-Hispanic Whites uses the model based on the Wave 2 core

sample while separate Wave 2 models are used for Asians, AIANs, and Native Hawaiians. The estimation for the remaining races, including multiple races, uses the overall average of the probability of increase for Asians, AIANs, and Native Hawaiians. For further details regarding the models, see Mulry and Keller (2007).

For Hispanics, Non-Hispanic Whites, and Non-Hispanic Blacks, the models used the average awareness estimated within subgroups defined by language spoken at home with all race and ethnicity groups pooled. For Asians, Native Hawaiians, and AIAN, the average awareness was the one measured for each group.

Overall, the approach estimates an increase of 5.5 million mail returns in all eight clusters combined due to the PMP. Since there are 102.8 million occupied housing units in the mailout/mailback areas in the eight clusters, this implies that PMP increased the mail return rate by about 5.3 percentage points in these areas.

These results are corroborated somewhat by the increase in mail response observed by a nationwide test of the American Community Survey (ACS). During the months January through March, the mail response was 5 to 9 percentage points higher in 2000 than in those same months in 2001 when there was no advertising by the Census Bureau (Bentley, Trancreto, and Hill 2006).

Tables 6 and 7 show how the results of the logistic regression models may be used to classify the clusters by high, medium, and low for the increase in probability of response

due to advertising and marketing, and for the increase in number of responses due to advertising and marketing.

The clusters with the largest increase in probability of mail return due to PMP are not necessarily the clusters that will produce the largest increase in number of mail returns. For the larger groups a small increase in the probability of mailing back a census form produces a large number of forms. For example, the estimated percentage increase for the Advantage Homeowner cluster with almost 27 million occupied households is the lowest at 4.08 percent, but the estimated increase in number of mail returns is over one million.

Summary and Discussion

In summary, the groups emerging from the cluster analysis present contrasting socioeconomic and demographic pictures according to propensity to mail back a census form in 2000. It is interesting to note that some of the clusters have very similar mail return rates and HTC scores yet look very different once we more closely examine the characteristics that compose the tracts – this is the type of detail that should help inform the communications contractor as they develop tailored media messages and delivery strategies.

It is also of interest that the clusters mirror in many ways the "stairstep" typology of household characteristics correlated with mail return documented by Word (1997). Word

noted that in the 1990 Census, White, non-Hispanic owners in spousal households had the lowest non-mailback rate (13.2%) while Hispanic renters in non-spousal households had the highest non-mailback rate (64.3%). In keeping with this typology, our highest mail return cluster (the Advantaged Homeowners) had the highest percentage White population, lowest percentage of renters, and lowest percent of non-spousal households. In contrast, the cluster with the lowest mail return rate (Economically Disadvantaged II – Renter Skewed) had the lowest percentage of Whites, highest percentage of renters, and highest percent of non-spousal households.

Theoretically, our results can also be taken down to a specific geography if the goal is to target a particular Census tract. However, several limitations and caveats are warranted. First, users must realize that the PDB reflects characteristics of tracts as they were in 2000. Urban renewal, gentrification, natural disasters and other factors can lead to drastic tract changes since 2000. Additionally, the physical boundaries for some tabulation tracts change over the decade by expanding or shrinking. Therefore, a tract falling into a particular cluster in 2000 may no longer be a good representative of that cluster seven years later.

Consequently, we recommend that multiple indicators be used to understand how best to market to a given tract. Specifically, we recommend looking together at the tract HTC score, MRR, cluster number, and whether the tract scored high on any of the three factor analysis scores. Using any of these indicators in isolation could be misleading. For example, as a result of successful partnership activities and committed community and

local leaders, a tract with a high HTC score that scores high on Factor 3 and falls into Cluster 6 might have had an unexpectedly high MRR in 2000 – using only the MRR to characterize the tract would be misleading in this case. Although the additional efforts of local officials in some areas produced a MRR higher than observed in other areas with apparently similar characteristics, quantifying the effect and incorporating it into statistical models is not possible at this time.

Estimating the increase in mail return rate using the logistic regression models provides a way of estimating the gain in mail response due to the PMP by population cluster. The observed mail return rates for the clusters alone are not sufficient for gauging the impact since they were influenced by the PMP. Potentially, the estimates of the increase in response rate and in the number of responses by segment will aid in achieving the goals of increasing the overall mail response rate and improving the 2010 Census count for the hard-to-count segments of the population. Even a small increase in the mail response rate in a large cluster that tends to have a high response rate will increase the overall response rate and reduce costs. Increases in the mail response rate in the smaller hard-to-count segments will aid in reducing the differential undercount observed for these groups in previous censuses.

The segmentation scheme presented in this report was prepared with the 2010 Integrated Communications Program (ICP) contractor in mind. We are hopeful it may serve as a foundation upon which to allocate resources, build Census 2010 market campaign strategies and ultimately, develop a larger communications plan. In the future, we hope

our plan can be supplemented with traditional marketing databases to help further describe the clusters in terms of consumer psychographics, media consumption habits and the like. We also hope to use Census tests (e.g., the 2006 Census Test and 2008 Dress Rehearsal) to further validate, test, and refine the model. Finally, as mentioned in the introduction, this research is a macro-level analysis designed to form a strategy for segmenting the mailback population. Additional micro-level analysis using ACS data, Census 2000 100% Detail files and Census 2000 survey data are planned to supplement the macro analysis.

Acknowledgments: The authors wish to express their gratitude to the Dallas Regional Office geography section for producing the cluster map. Thanks also to Marie Pees for providing various decennial Census data files necessary for the cluster and mail return rate analyses. Finally, thanks to our reviewers Pat Cantwell, Aref Dajani, and Gregg Robinson for very helpful comments on earlier versions.

References

Baron, S. and Billia, D. (1999). Building a Surrogate for Predicting Census Participation." A paper presented at the 54th Annual Conference for the American Association for Public Opinion Research, St. Petersberg, Florida, 1999.

Bates N., Dahlhamer, J. and Singer, E. (2006). "Privacy Concerns, Too Busy or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse." A poster presented at the 2006 annual conference of the American Association for Public Opinion Research.

Bates N. and Piani, A. (2005). "Participation in the National Health Interview Survey: Exploring Reasons for Reluctance Using Contact History Process Data." *Proceedings of the Federal Committee on Statistical Methods Research Conference.* Office of Management and Budget. Washington, DC. http://www.fcsm.gov/05papers/Bates_Piani_VIB.pdf

Bentley, M., Tancreto, J., and Hill, J. (2006) "Research on Census Response Rates and Partnerships and Marketing". Internal report dated September 6, 2006. Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Bruce, A. (2003). "Application and Evaluation of the Planning Database for 1990 and 2000: Comparability Measures of Mail Return Rates and Potential Use in Census 2010 Planning." U.S. Census Bureau, unpublished report, August 4, 2003.

Hilligus, D.S., Nie N., Prewitt K. and Pals, H. (2006). The Hard Count: The Political and Social Challenges of Census Mobilization. The Russell Sage Foundation: New York: New York.

Mulry, M. H. and Keller, J. K. (2007) "Investigation of the Relationship between Census Mail Response and the Marketing and Partnership Campaign". *2007 Joint Statistical Meeting Proceedings, American Association of Public Opinion Research* [CD-ROM]. American Statistical Association. Alexandria, VA. (to appear)

Robinson, J.G., Johanson, C. and Bruce, A. (2007). "The Planning Database: Decennial Data for Historical, Real-time, and Prospective Analysis", paper presented at the 2007 Joint Statistical Meetings, Salt Lake City, Utah.

SAS® Procedures Guide, Version 8. SAS Institute, Inc.: Cary, N.C.

U.S. Census Bureau (2006). "2010 Census Communications Campaign Contract." Solicitation No. YA-1323-07-FR-002.

U.S. Census Bureau (2001). "Technical Documentation: Census 2000 Summary File 1, Census 2000 Geographic Terms and Concepts", pp. A-22.

Wolter, K. et al (2003) "Partnership and Marketing Program Evaluation". Census 2000 Evaluation D.1. Prepared by NORC. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/D.1.pdf

Word, D.L. (1997). "Who Responds/Who Doesn't? Analyzing Variation in Mail Response Rates During the 1990 Census." U.S. Census Bureau, Population Division Working Paper no. 19, July 1997.

Table 1. Factor Analysis of Census 2000 Planning Database Hard-to-Count Variables

	Factor 1 – ***Economically disadvantaged***	Factor 2 – ***Unattached/mobile singles***	Factor 3 – ***High density w/ethnic enclaves***
Underlying housing and social characteristics:	- Vacant housing - Poverty - Public Assistance - Unemployment - Less than high school education - No phone	- Multiunit structures - Renters - Nonspousal households - Persons moved in last year	- Crowded housing - Few vacant houses - Linguistic isolation - Less than high school education
Correlated demographic characteristics:	- High correlation with % black - Moderate correlation with % AIAN - Moderate correlation with % pop <18 (children)	- No strong correlation with any one race/ethnicity (racially/ethnically diverse) - Moderate *negative* correlation with % pop <18 (absence of children)	- High correlation with % Hispanic - Moderate correlation with % Asian or NHPI - Moderate correlation with % pop <18 (children)
% variance explained by each factor: (cumulative =74.7%)	46.2%	14.7%	13.8%
Average 2000 mail return rate for tracts with high factor score: (national avg. MRR=75.4%)	63.5%	66.5%	67.2%
Average HTC score for tracts with high factor score: (national avg. HTC score=33)	75	65	75
Pearson correlation coefficient with **2000 MRR**	-.56	-.48	-.21
Number of tracts and % of total tracts **w/ high factor score**	N=7051 (11.2%)	N=4073 (6.5%)	N=3758 (6.0)%

Table 2. Mail Return Rate, HTC Score, Number of Tracts and Occupied Housing Units by Cluster

#	Cluster Name	Mail Return Rate	HTC Score	Total Occupied Housing Units		Number of Tracts
				Number (in millions)	Percent	
1	All around average I (homeowner skewed)	77.3%	23	36.5	35%	21,174
2	All around average II (renter skewed)	74.2%	41	16.5	16%	8,957
3	Econ. Disadvantaged I (homeowner skewed)	66.5%	65	6.6	6%	5,230
4	Econ. Disadvantaged II (renter skewed)	58.0%	92	3.0	3%	2,574
5	Ethnic Enclave I (homeowner skewed)	69.8%	63	3.4	3%	2,440
6	Ethnic Enclave II (renter skewed)	63.6%	84	2.5	2%	1,754
7	Young/mobile/singles	67.1%	61	8.0	8%	4,073
8	Advantaged Homeowners	83.2%	6	26.8	26%	16,506

Table 3. Race and Ethnicity Population Distribution[1] Among the Clusters

Cluster number	Cluster name	Non-Hisp. White		Black		Hispanic		Asian		AIAN		NHPI	
		%	number	%	number	%	number	%	number	%	number	%	number
1	All Avg. I	39.9	75,485	25.4	8,633	19.0	6,620	20.8	2,101	35.6	825	29.5	115
2	All Avg. II	14.6	27,689	13.8	4,697	14.0	4,857	19.6	1,977	14.0	324	19.6	76
3	Econ. Dis. I	3.6	6,797	25.3	8,569	5.3	1,825	2.5	250	15.0	347	5.2	20
4	Econ. Dis. II	0.8	1,597	11.9	4,045	5.6	1,956	2.3	234	2.5	58	2.6	10
5	Ethnic Encl. I	1.4	2,636	2.8	961	22.0	7,661	5.7	571	12.2	283	8.1	32
6	Ethnic Encl. II	0.8	1,514	2.3	779	14.8	5,129	8.8	886	3.4	78	5.9	23
7	Single/mobile	5.3	9,960	8.3	2,809	7.1	2,473	13.1	1,315	5.0	116	9.1	36
8	Advant. Homeowner	33.6	63,697	10.1	3,440	12.2	4,237	27.2	2,743	12.4	288	20.0	78
	Total	100.0	189,375	100.0	33,935	100.0	34,759	100.0	10,077	100.0	2,319	100.0	389

[1] numbers are in thousands

Table 4. Tracts with High Factor Scores by Cluster

#	Cluster Name	High Score[1] on Factor 1 ***Econ. Disadvantaged***	High Score[2] on Factor 2 ***Unattached/mobile/singles***	High Score[3] on Factor 3 ***High density/ethnic enclaves***
1	All around average (homeowner skew)	**16.5%** (1160)	**0%** (2)	**1.5%** (55)
2	All around average (renter skewed)	**1.4%** (100)	**9.9%** (403)	**3.8%** (144)
3	Econ. Disadvantaged (homeowner skewed)	**41.8%** (2944)	**0%** (6)	**1.0%** (37)
4	Econ. Disadvantaged (renter skewed)	**24.4%** (1722)	**1.4%** (57)	**8.6%** (323)
5	Ethnic Enclave (homeowner skewed)	**11.0%** (778)	**0%** (0)	**39.1%** (1468)
6	Ethnic Enclave (renter skewed)	**3.8%** (266)	**17.2** (702	**42.0%** (1580)
7	Young/mobile/singles	**1%** (61)	**71.2%** (2900)	**3.0%** (113)
8	Advantaged Homeowners	**0%** (20)	**0%** (3)	**1.0%** (38)
	Total tracts	**100%** (7,051)	**100%** (4,073)	**100%** (3,758)

[1] factor score >=1.25
[2] factor score >=1.75
[3] factor score >=1.75

Table 5. Estimated increase in mailing back a census form due to PMP by cluster based on Wave 2 models using the average awareness of PMP from Wave 3

#	Cluster Name	Occupied housing units in mailout/mailback universe (thousands)	Estimated increase due to PMP (thousands)	Estimated percentage increase due to PMP
1	All around avg. I	36,201	1,830	5.05%
2	All around avg. II	16,509	948	5.74%
3	Econ. Disadv. I	6,573	475	7.23%
4	Econ. Disadv. II	2,976	245	8.24%
5	Ethnic enclave I	3,249	218	6.69%
6	Ethnic enclave II	2,541	168	6.62%
7	Young/mobile/single	8,018	483	6.03%
8	Advan. Homeowner	26,753	1,092	4.08%
Total		102,820	5,459	5.31%

Table 6. Clusters classified by estimated percentage increase in Census 2000 mail return rate due to PMP

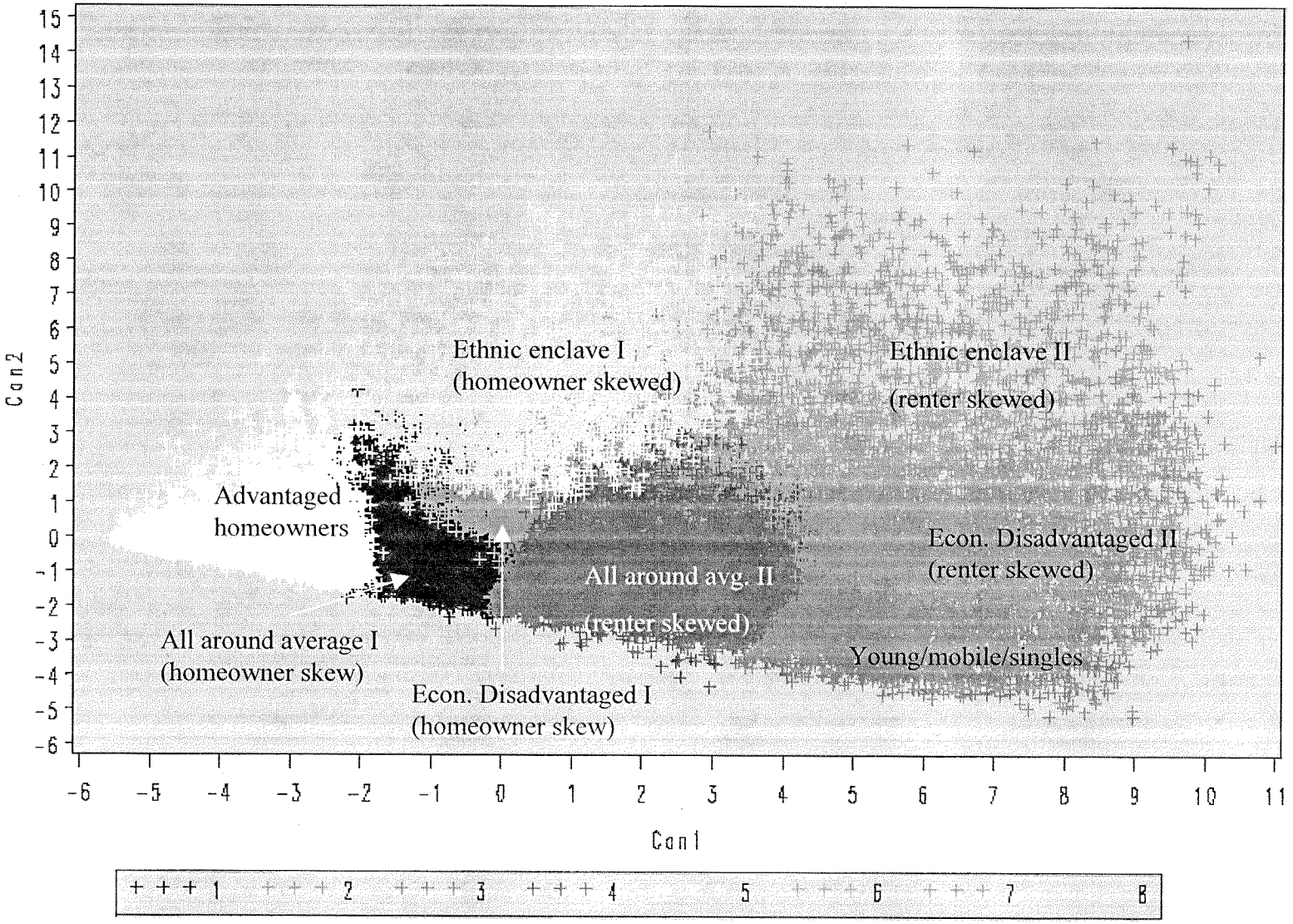
Very high (>= 8%)	**High (< 8% and >= 7%)**	**Medium (< 7% and >= 5.3%)**	**Low (< 5.3%)**
Cluster 4 (Econ. Disadv. II)	Cluster 3 (Econ. Disadv. I)	Cluster 7 (Young/mobile/single) Cluster 2 (All avg. II) Cluster 5 (Ethnic enclave I) Cluster 6 (Ethnic enclave II)	Cluster 1 (All avg. I) Cluster 8 (Advan. Homeowners)

Table 7. Clusters classified by estimated increase in numbers of occupied housing units mailing back a Census 2000 form due to PMP

Very high (>=1.5 mil)	**High (<1.5 mil and >=400,00)**	**Medium (<400,000 and >=175,000)**	**Low (<175,000)**
Cluster 1 (All avg. I)	Cluster 2 (All avg. II) Cluster 8 (Advan. Homeowners)	Cluster 3 (Econ. Disadv. I) Cluster 7 (Young/mobile/single) Cluster 4 (Econ. Disadv. II) Cluster 5 (Ethnic enclave I)	Cluster 6 (Ethnic enclave II)

cluster analysis

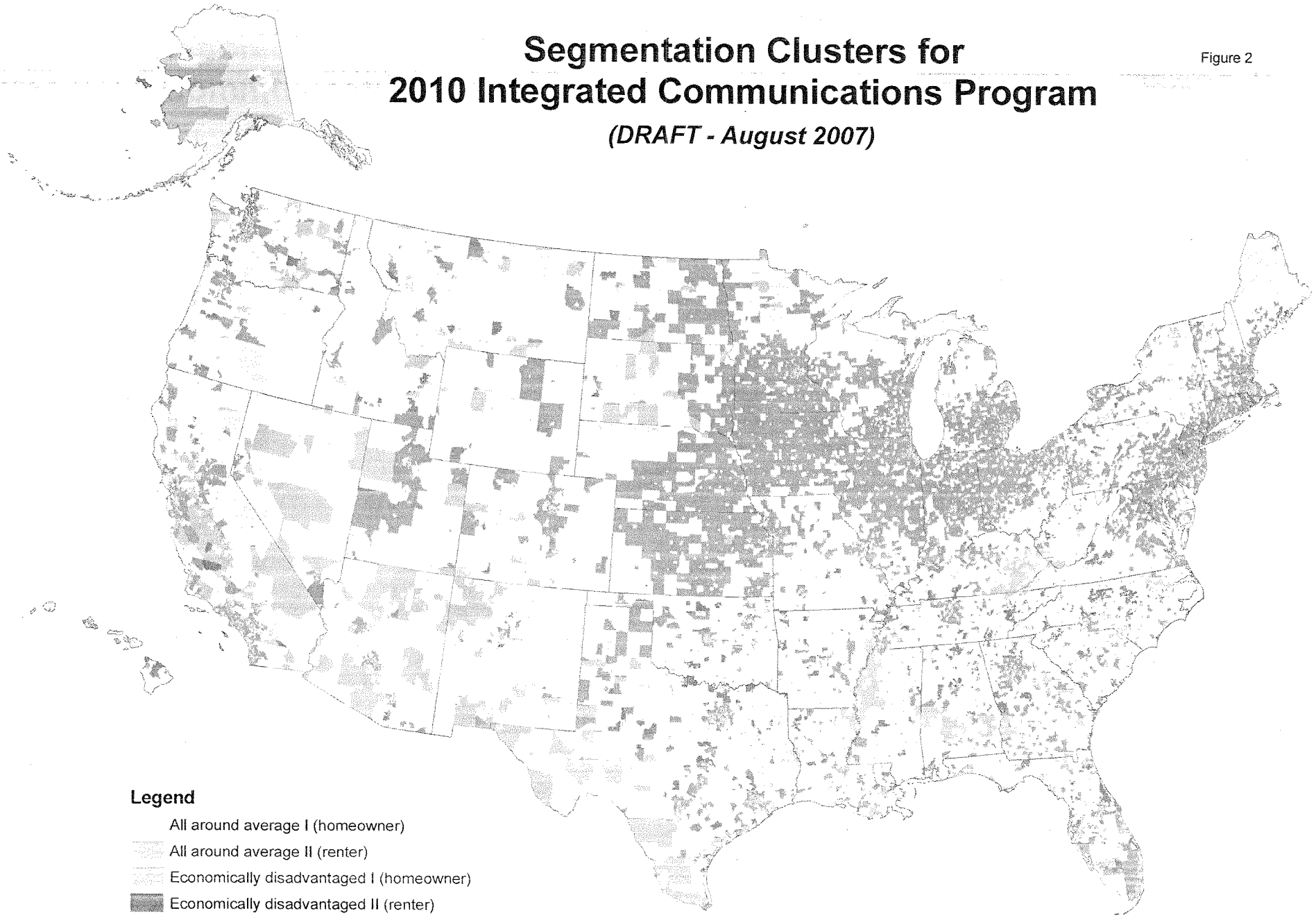
Figure 1.



Segmentation Clusters for 2010 Integrated Communications Program

(DRAFT - August 2007)

Figure 2



Source: Tract-Level Planning Database with Census 2000 Data
August 29, 2007

The FACTOR Procedure

Means and Standard Deviations from 62707 Observations

Variable	Mean	Std Dev
Pct_Vacant	7.626131	6.024354
Pct_Not_sing_unit	32.930861	24.251133
Pct_rent	34.231269	22.688604
Pct_crowd	2.937582	5.882433
Pct_poverty	13.289672	11.216721
Pct_PubAssis	3.897126	4.495376
Pct_unemploy	3.791641	2.746422
Pct_NotHusbWife	48.810056	16.009749
Pct_LIH	4.295037	7.383027
Pct_NO_HS	20.819233	13.906509
Pct_moved	19.307535	9.457424
Pct_NoPhone	2.775997	3.715250

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 12 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	5.53982183	3.77152482	0.4617	0.4617
2	1.76829701	0.11608309	0.1474	0.6090
3	1.65221393	0.90250998	0.1377	0.7467
4	0.74970394	0.22704386	0.0625	0.8092
5	0.52266009	0.04978575	0.0436	0.8527
6	0.47287434	0.10999096	0.0394	0.8921
7	0.36288337	0.07639544	0.0302	0.9224
8	0.28648793	0.07295594	0.0239	0.9462
9	0.21353199	0.03259832	0.0178	0.9640
10	0.18093367	0.02829897	0.0151	0.9791
11	0.15263470	0.05467750	0.0127	0.9918
12	0.09795720		0.0082	1.0000

3 factors will be retained by the NFACTOR

Factor Pattern

	Factor1	Factor2	Factor3
Pct_Vacant	35	-28	60 *
Pct_Not_sing_unit	69 *	54 *	4
Pct_rent	82 *	50 *	-1
Pct_crowd	60 *	-9	-69 *
Pct_poverty	88 *	-21	16
Pct_PubAssis	77 *	-31	3
Pct_unemploy	70 *	-22	12
Pct_NotHusbWife	74 *	37 *	33
Pct_LIH	57 *	0	-72 *
Pct_NO_HS	76 *	-46 *	-19
Pct_moved	45 *	66 *	12
Pct_NoPhone	63 *	-43 *	31

Printed values are multiplied by 100 and rounded to the nearest Values greater than 0.35 are flagged by an '*'.

The FACTOR Procedure
Initial Factor Method: Principal Components

Variance Explained by Each Factor

Factor1	Factor2	Factor3
5.5398218	1.7682970	1.6522139

Final Communality Estimates: Total = 8.960333

Pct_Vacant	Pct_Not_sing_unit	Pct_rent	Pct_crowd	Pct_poverty	Pct_PubAssis
0.55630120	0.76349636	0.91745422	0.83741053	0.84802607	0.69124378

Pct_unemploy	Pct_NotHusbWife	Pct_LIH	Pct_NO_HS	Pct_moved	Pct_NoPhone
0.55803647	0.78581369	0.84851200	0.82036084	0.65226966	0.68140795

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2	3
1	0.69500	0.57210	0.43553
2	-0.59056	0.79972	-0.10809
3	0.41014	0.18209	-0.89366

Rotated Factor Pattern

	Factor1	Factor2	Factor3
Pct_Vacant	65 *	9	-36 *
Pct_Not_sing_unit	18	83 *	20
Pct_rent	27	86 *	31
Pct_crowd	19	14	88 *
Pct_poverty	80 *	37 *	26
Pct_PubAssis	73 *	20	34
Pct_unemploy	67 *	24	22
Pct_NotHusbWife	43 *	77 *	-1
Pct_LIH	10	20	89 *
Pct_NO_HS	72 *	3	55 *
Pct_moved	-3	81 *	2
Pct_NoPhone	82 *	7	5

Printed values are multiplied by 100 and rounded to the nearest Values greater than 0.35 are flagged by an '*'.

Variance Explained by Each Factor

Factor1	Factor2	Factor3
3.5704912	2.9988417	2.3909999

Final Communality Estimates: Total = 8.960333

Pct_Vacant	Pct_Not_sing_unit	Pct_rent	Pct_crowd	Pct_poverty	Pct_PubAssis
0.55630120	0.76349636	0.91745422	0.83741053	0.84802607	0.69124378

The FACTOR Procedure
Rotation Method: Varimax

Pct_unemploy	Pct_NotHusbWife	Pct_LIH	Pct_NO_HS	Pct_moved	Pct_NoPhone
0.55803647	0.78581369	0.84851200	0.82036084	0.65226966	0.68140795

The FACTOR Procedure
Rotation Method: Varimax

Scoring Coefficients Estimated by Regression

Squared Multiple Correlations of the Variables with Each Factor

Factor1	Factor2	Factor3
1.0000000	1.0000000	1.0000000

Standardized Scoring Coefficients

	Factor1	Factor2	Factor3
Pct_Vacant	0.28461	-0.02292	-0.28065
Pct_Not_sing_unit	-0.08243	0.31822	-0.00108
Pct_rent	-0.06682	0.30869	0.04007
Pct_crowd	-0.06573	-0.05566	0.42490
Pct_poverty	0.21893	0.01602	-0.00425
Pct_PubAssis	0.20611	-0.05526	0.06394
Pct_unemploy	0.19230	-0.01548	0.00477
Pct_NotHusbWife	0.05218	0.27811	-0.14336
Pct_LIH	-0.10672	-0.02084	0.43520
Pct_NO_HS	0.20083	-0.14889	0.18973
Pct_moved	-0.13363	0.35790	-0.06993
Pct_NoPhone	0.29993	-0.09515	-0.09134

Cluster analysis using 12 HTC variables: 2000 Tract-Level (maxiter=100)

Cluster Summary

Cluster	Frequency	RMS Std Deviation	from Seed to Observation	Radius Exceeded	Nearest Cluster	Distance Between Cluster Centroids
1	21174	5.9484	64.5845		8	28.3450
2	8957	7.2178	98.7144		3	35.3832
3	5230	7.3918	92.4019		2	35.3832
4	2574	9.6140	106.8		7	38.6577
5	2440	9.0242	110.2		3	35.6263
6	1754	10.406	99.2514		4	42.4771
7	4073	8.5291	93.4796		4	38.6577
8	16506	4.7549	88.4856		1	28.3450

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=8 Maxiter=100 Converge=0.02

Statistics for Variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
Pct_Vacant	6.0251	5.53535	0.156056	0.184912
Pct_Not_sing_unit	24.25112	10.68225	0.805995	4.154498
Pct_rent	22.68846	9.21614	0.835017	5.061221
Pct_crowd	5.88243	3.91649	0.556767	1.256149
Pct_poverty	11.21672	6.70388	0.642832	1.799802
Pct_PubAssis	4.49538	3.19174	0.495949	0.983926
Pct_unemploy	2.74642	2.19367	0.362088	0.567614
Pct_NotHusbWife	16.00968	8.45019	0.721440	2.589890
Pct_LIH	7.38303	4.59831	0.612137	1.578227
Pct_NO_HS	13.90651	8.58173	0.619228	1.626246
Pct_moved	9.45742	7.41104	0.385946	0.628521
Pct_NoPhone	3.71525	3.03933	0.330838	0.494407
OVER-ALL	12.69784	6.67915	0.723348	2.614644

Pseudo F Statistic = 23419.74

Approximate Expected Over-All R-Squared = 0.54672

Cubic Clustering Criterion = 402.581

Cluster Means

Cluster	Pct_Vacant	Not_sing_unit	Pct_rent	Pct_crowd	Pct_poverty	Pct_PubAssis
1	8.78605026	27.76241	25.05825	1.2680413	10.7653107	2.85997483
2	6.81471607	49.64294	48.21095	2.5480391	11.5151426	3.20120715
3	12.18270977	29.29975	46.33612	3.0167264	28.2651813	8.67369770
4	12.43299209	79.67669	81.14152	5.9199271	38.9874317	14.19206922
5	7.44187569	25.35898	37.24450	15.115105	24.8246506	8.11467756
6	5.51413362	73.16869	74.66962	22.516099	28.2676234	9.69574373
7	7.52254739	79.93038	74.02117	4.0075231	16.9379530	3.42125585
8	4.66383212	9.597491	12.69989	0.6557123	4.54088282	1.36395113

Cluster	Pct_unemploy	Pct_NotHusb Wife	Pct_LIH	Pct_NO_HS	Pct_moved	Pct_NoPhone
1	3.22026557	45.28217342	1.9344417	20.0105905	16.8837152	2.65828808
2	3.57520153	57.28353873	4.2932585	16.4282898	25.6691445	1.77523255
3	6.55233467	67.00361221	2.9312485	34.7904927	20.6474364	7.0174283
4	8.88017192	79.26376448	8.6349345	40.0626820	25.3354396	8.46825732
5	5.96344651	44.25791775	18.416104	49.8394369	18.5674208	5.69024188
6	6.10453439	54.18542781	31.266865	50.0700423	24.7436955	4.9605505
7	4.16228509	71.82124944	6.7947428	15.3986933	34.5124983	2.1335321
8	2.31542161	32.64656533	1.5089463	10.7502948	13.3795446	0.73394573

Cluster Standard Deviations

Cluster	Pct_Vacant	Not_sing_unit	Pct_rent	Pct_crowd	Pct_poverty	Pct_PubAssis
1	6.45124856	9.90697	7.86309	1.8523384	5.5455932	2.03524996
2	4.82758290	12.10127	11.51268	3.1418019	6.1007114	2.61137276
3	5.36949400	11.86706	10.91631	3.3143951	9.2444078	5.06790637
4	7.13279831	15.34751	11.56861	5.2888154	12.320900	8.03470272
5	5.68927160	13.13121	13.22423	10.301941	9.9991769	5.07480459
6	3.63948028	17.30463	12.33019	13.557607	10.347102	6.01884252
7	5.04611966	12.56199	11.72799	3.7946103	10.218763	3.01526992
8	4.56334138	7.111046	6.07352	1.3829933	2.9035229	1.26011842

Cluster	Pct_unemploy	Pct_NotHusb Wife	Pct_LIH	Pct_NO_HS	Pct_moved	Pct_NoPhone
1	1.68629052	7.59149090	2.781121	8.98142387	5.3717495	2.63524522
2	2.14634456	9.63982274	4.583055	8.52291215	8.6053327	1.93012166
3	3.12208960	8.54947787	4.146172	8.81979717	6.7143213	4.67030748
4	4.58690149	8.21257450	8.762619	10.1527586	11.860822	5.96083392
5	2.92596772	8.13263167	9.535136	10.8416937	6.0321806	6.79323533
6	2.52196382	8.88780543	11.46399	13.7080211	9.0335758	4.61417169

7	2.72465592	10.9654638	5.943999	8.84099968	13.326823	2.33706344
8	1.41168969	8.08735639	2.340353	6.28860382	6.1148571	1.13306672

cluster_new=AVERAGE I HMOWN

The MEANS Procedure

Variable	Sum	Mean	Minimum	Maximum	Median
tot_pop	95000325.00	4486.65	110.0000000	24506.00	4192.00
total_HU	40021542.00	1890.13	56.0000000	9118.00	1779.00
tot_occ_HU	36469966.00	1722.39	47.0000000	8574.00	1623.00
mail_RR	1624669.11	77.3173326	20.3517588	100.0000000	78.0667702
HTC	485333.00	22.9211769	0	81.0000000	21.0000000
pct_urban	1337514.02	63.1677537	0	100.0000000	90.2374030
pct_rural	779885.98	36.8322463	0	100.0000000	9.7625970
avg_hhd_sz	54069.46	2.5535780	1.4200000	5.0500000	2.5300000
pop_sq_mile	43492458.70	2054.05	0.0327114	44892.87	559.0135499
Pct_Black	192784.46	9.1047730	0	99.0016639	2.0261681
Pct_Hisp	138729.52	6.5518807	0	90.8682214	2.2662868
Pct_nonHisp_wht	1693094.88	79.9610316	0	99.6757633	87.9354390
Pct_Asian	42764.16	2.0196543	0	76.5927589	0.5437651
pct_nhpi	2320.43	0.1095887	0	46.6880654	0.0200622
pct_api	45084.59	2.1292430	0	80.8382482	0.5904785
pct_AIAN	20692.13	0.9772426	0	99.1317671	0.3290228
pct_LIHH_span	21503.97	1.0155838	0	29.6901408	0.2916768
Pct_sing_units	1529558.77	72.2375917	15.5649626	100.0000000	72.8095879
Pct_10_units	115411.09	5.4506039	0	70.3221385	2.4299102
pct_mob_home	277179.35	13.0905523	0	84.1745360	8.2368965
pct_gq_noninst	13922.75	0.6575397	0	47.6016182	0.0635055
Pct_65plus	309660.48	14.6245622	1.1771117	89.2814473	13.8209347
pct_pop_lt18	527641.96	24.9193333	0	50.6349462	25.0678279

cluster_new=AVERAGE II RENT

Variable	Sum	Mean	Minimum	Maximum	Median
tot_pop	40394075.00	4509.78	117.0000000	36146.00	4254.00
total_HU	17780883.00	1985.14	56.0000000	9757.00	1870.00
tot_occ_HU	16535594.00	1846.11	53.0000000	9319.00	1751.00
mail_RR	663053.51	74.1836550	1.1006289	99.8491704	75.0610619
HTC	366424.00	40.9092330	9.0000000	94.0000000	39.0000000
pct_urban	871768.64	97.3281949	0	100.0000000	100.0000000
pct_rural	23931.36	2.6718051	0	100.0000000	0
avg_hhd_sz	21510.77	2.4015597	1.2900000	4.6600000	2.3500000
pop_sq_mile	50579173.99	5646.89	0.1353604	121575.78	3899.48
Pct_Black	105584.11	11.7878872	0	98.9298454	4.5628141
Pct_Hisp	101007.65	11.2769513	0.0730994	94.2491364	6.2317997
Pct_nonHisp_wht	621477.38	69.3845460	0	99.2184592	75.6340838
Pct_Asian	41437.05	4.6262201	0	71.6177179	2.2847949
pct_nhpi	1580.17	0.1764172	0	30.8377309	0.0479042

cluster_new=AVERAGE II RENT

The MEANS Procedure

Variable	Sum	Mean	Minimum	Maximum	Median
pct_api	43017.22	4.8026373	0	79.0437436	2.4053975
pct_AIAN	7465.05	0.8334319	0	95.6185567	0.3797949
pct_LIHH_span	18212.06	2.0332768	0	32.2854448	0.8849558
Pct_sing_units	451048.16	50.3570568	0	100.0000000	51.6243655
Pct_10_units	178132.35	19.8875017	0	99.4663109	18.0842607
pct_mob_home	45828.48	5.1164987	0	99.0583804	0.2671756
pct_gq_noninst	12861.98	1.4359695	0	49.9737257	0.1566989
Pct_65plus	120221.90	13.4221167	0	96.7692308	12.2048451
pct_pop_lt18	205986.42	22.9972554	0	53.7001898	23.0031949

cluster_new=Econ Disad I HMOWN

Variable	Sum	Mean	Minimum	Maximum	Median
tot_pop	18021065.00	3445.71	104.0000000	12100.00	3217.00
total_HU	7528035.00	1439.39	55.0000000	5227.00	1358.00
tot_occ_HU	6644402.00	1270.44	51.0000000	4847.00	1193.00
mail_RR	341899.38	66.5044512	0.6802721	100.0000000	66.8730650
HTC	341532.00	65.3024857	22.0000000	106.0000000	65.0000000
pct_urban	480027.42	91.7834448	0	100.0000000	100.0000000
pct_rural	42972.58	8.2165552	0	100.0000000	0
avg_hhd_sz	13864.45	2.6509465	1.1400000	4.6200000	2.6100000
pop_sq_mile	24938821.23	4768.42	0.0626146	50904.70	3239.48
Pct_Black	256621.25	49.0671599	0	99.5052226	48.9009340
Pct_Hisp	48452.37	9.2643150	0	91.4077417	2.2948037
Pct_nonHisp_wht	192732.95	36.8514245	0	98.9230156	33.0919141
Pct_Asian	6427.15	1.2288997	0	52.2135766	0.3522678
pct_nhpi	499.8646313	0.0955764	0	29.6686747	0
pct_api	6927.01	1.3244762	0	69.4928897	0.3949849
pct_AIAN	12005.51	2.2955095	0	97.5323149	0.3111182
pct_LIHH_span	10661.30	2.0384891	0	29.8013245	0.5499048
Pct_sing_units	369762.29	70.7002466	33.1835206	100.0000000	70.5031659
Pct_10_units	32851.68	6.2813912	0	52.5452977	3.9315697
pct_mob_home	25363.70	4.8496557	0	60.8695652	0.7159306
pct_gq_noninst	6601.56	1.2622481	0	47.5713756	0.1811349
Pct_65plus	66977.04	12.8063173	2.0864382	61.5658363	12.4118753
pct_pop_lt18	152427.37	29.1448134	5.5555556	51.4688602	29.1043551

cluster_new=Econ Disad II RENT

The MEANS Procedure

Variable	Sum	Mean	Minimum	Maximum	Median
tot_pop	7930558.00	3081.02	135.0000000	15838.00	2677.00
total_HU	3359772.00	1305.27	67.0000000	6660.00	1181.50
tot_occ_HU	2970176.00	1153.91	48.0000000	6408.00	1037.00
mail_RR	149175.26	57.9997129	18.1818182	86.6666667	58.1182136
HTC	237173.00	92.1418026	46.0000000	125.0000000	93.0000000
pct_urban	257083.38	99.8769917	0	100.0000000	100.0000000
pct_rural	316.6232463	0.1230083	0	100.0000000	0
avg_hhd_sz	6571.10	2.5528749	1.0800000	4.3400000	2.5700000
pop_sq_mile	55915540.46	21723.21	11.1274693	229694.24	10435.58
Pct_Black	137656.12	53.4794546	0	100.0000000	53.5407225
Pct_Hisp	52931.63	20.5639581	0	96.6232513	9.6661305
Pct_nonHisp_wht	56137.43	21.8094116	0	93.8154139	9.9969615
Pct_Asian	6648.19	2.5828229	0	62.2341669	0.7406140
pct_nhpi	371.2222859	0.1442200	0	63.3237822	0.0292719
pct_api	7019.41	2.7270429	0	79.0830946	0.8247318
pct_AIAN	1959.06	0.7610958	0	97.9899497	0.4310345
pct_LIHH_span	15616.40	6.0669756	0	41.0579345	2.6190263
Pct_sing_units	52312.19	20.3233051	0	75.2873563	17.4519101
Pct_10_units	97955.96	38.0559294	0	100.0000000	31.6429894
pct_mob_home	1800.73	0.6995862	0	88.4536082	0
pct_gq_noninst	8439.95	3.2789239	0	49.4230428	0.5051665
Pct_65plus	27691.26	10.7580633	0	61.9647355	9.1867221
pct_pop_lt18	74725.47	29.0308746	0.2690397	58.7262200	30.4525981

cluster_new=Ethnic I HOMEOWN

Variable	Sum	Mean	Minimum	Maximum	Median
tot_pop	12173442.00	4989.12	126.0000000	24441.00	4749.00
total_HU	3638196.00	1491.06	54.0000000	7748.00	1389.00
tot_occ_HU	3361078.00	1377.49	48.0000000	6691.00	1295.00
mail_RR	168038.26	69.7833312	33.2878581	87.9106439	70.2665321
HTC	153305.00	62.8299180	27.0000000	104.0000000	62.0000000
pct_urban	205715.95	84.3098176	0	100.0000000	100.0000000
pct_rural	38284.05	15.6901824	0	100.0000000	0
avg_hhd_sz	8746.96	3.5848197	2.2400000	6.7500000	3.5400000
pop_sq_mile	14643903.88	6001.60	0.0830327	47283.84	4726.28
Pct_Black	19300.58	7.9100747	0	87.7515614	2.1428862
Pct_Hisp	148041.51	60.6727495	0.0881834	99.2414248	64.5789685
Pct_nonHisp_wht	58541.12	23.9922640	0.2403846	99.5507301	15.4232711
Pct_Asian	11059.29	4.5324947	0	81.7215728	0.6866945
pct_nhpi	606.1318073	0.2484147	0	33.1555556	0.0577090

cluster_new=Ethnic I HOMEOWN

The MEANS Procedure

Variable	Sum	Mean	Minimum	Maximum	Median
pct_api	11665.42	4.7809093	0	87.4070138	0.7937096
pct_AIAN	5779.69	2.3687253	0	98.6865942	0.9194989
pct_LIHH_span	39151.51	16.0457001	0	56.0574082	14.8938853
Pct_sing_units	182124.09	74.6410197	27.5899673	100.0000000	75.5604818
Pct_10_units	14213.25	5.8251039	0	52.0113775	2.9465419
pct_mob_home	23110.57	9.4715442	0	69.0058480	2.2262156
pct_gq_noninst	1422.52	0.5829989	0	43.3675565	0.0650215
Pct_65plus	22475.57	9.2113004	2.0066890	30.1976823	8.4545086
pct_pop_lt18	80055.82	32.8097638	13.6316695	52.8011204	33.2946257

cluster_new=Ethnic II RENT

Variable	Sum	Mean	Minimum	Maximum	Median
tot_pop	8485915.00	4838.04	133.0000000	14475.00	4442.00
total_HU	2681712.00	1528.91	50.0000000	5599.00	1394.50
tot_occ_HU	2539704.00	1447.95	40.0000000	5302.00	1321.00
mail_RR	111595.56	63.6234684	24.3119266	86.4406780	64.2177834
HTC	146876.00	83.7377423	45.0000000	122.0000000	84.0000000
pct_urban	174853.64	99.6885090	0	100.0000000	100.0000000
pct_rural	546.3551609	0.3114910	0	100.0000000	0
avg_hhd_sz	5899.85	3.3636545	1.3900000	6.7700000	3.2900000
pop_sq_mile	50134880.40	28583.17	3.7858399	210550.96	20067.60
Pct_Black	15522.59	8.8498209	0	75.1319648	4.2769511
Pct_Hisp	103239.20	58.8592906	0.6544503	98.4146850	63.1821439
Pct_nonHisp_wht	33292.63	18.9809767	0.2724796	98.0377994	12.4361589
Pct_Asian	19308.57	11.0083041	0	95.0866142	3.9490935
pct_nhpi	513.6732941	0.2928582	0	42.8405122	0.0791139
pct_api	19822.24	11.3011623	0	95.0866142	4.1299791
pct_AIAN	1600.80	0.9126562	0	7.9264948	0.8051591
pct_LIHH_span	39204.63	22.3515548	0	78.9473684	21.1330994
Pct_sing_units	47062.12	26.8313124	0	81.1834320	24.9907203
Pct_10_units	56067.28	31.9653794	0	98.2945736	26.6809005
pct_mob_home	4241.31	2.4180806	0	88.2629108	0
pct_gq_noninst	1400.38	0.7983898	0	45.7521645	0.1113808
Pct_65plus	14363.05	8.1887394	0.7039644	42.2695035	6.8728720
pct_pop_lt18	53856.09	30.7047272	5.2216151	58.2272033	30.9724775

cluster_new=mobile/single

The MEANS Procedure

Variable	Sum	Mean	Minimum	Maximum	Median
tot_pop	17124794.00	4204.47	106.0000000	24523.00	3866.00
total_HU	8648255.00	2123.31	56.0000000	11522.00	1916.00
tot_occ_HU	8013178.00	1967.39	52.0000000	11168.00	1778.00
mail_RR	273601.57	67.1744595	32.2802198	88.8704319	67.8654292
HTC	246565.00	60.5364596	23.0000000	107.0000000	60.0000000
pct_urban	405863.66	99.6473505	0	100.0000000	100.0000000
pct_rural	1436.34	0.3526495	0	100.0000000	0
avg_hhd_sz	8533.76	2.0952026	1.1500000	4.6100000	2.0600000
pop_sq_mile	66960904.57	16440.19	1.9217546	201954.33	8298.89
Pct_Black	67717.72	16.6260052	0.0486855	99.0158666	7.6677316
Pct_Hisp	54327.96	13.3385605	0.2461538	78.7207872	8.9351285
Pct_nonHisp_wht	241648.48	59.3293590	0.1782884	98.6368062	63.7443439
Pct_Asian	30118.52	7.3946772	0	72.4005135	4.4559970
pct_nhpi	822.7116106	0.2019916	0	17.2254820	0.0653808
pct_api	30941.23	7.5966688	0	74.4607942	4.6197183
pct_AIAN	2812.98	0.6906401	0	25.2684776	0.3832190
pct_LIHH_span	11249.44	2.7619539	0	31.1142974	1.4371257
Pct_sing_units	81743.58	20.0696232	0	83.9080460	20.0527704
Pct_10_units	181022.60	44.4445369	0	100.0000000	42.2205990
pct_mob_home	6376.45	1.5655421	0	87.2384937	0
pct_gq_noninst	10246.60	2.5157389	0	49.7204867	0.2247191
Pct_65plus	43471.23	10.6730257	0	86.8571429	9.3719171
pct_pop_lt18	70332.86	17.2680717	0.3171118	49.0028490	17.2817282

cluster_new=Advd homeowner

Variable	Sum	Mean	Minimum	Maximum	Median
tot_pop	75428141.00	4569.74	101.0000000	34055.00	4223.00
total_HU	28132035.00	1704.35	51.0000000	10865.00	1579.50
tot_occ_HU	26782498.00	1622.59	44.0000000	10102.00	1507.00
mail_RR	1371814.95	83.2159510	4.3310131	100.0000000	83.9339991
HTC	99392.00	6.0215679	0	49.0000000	4.0000000
pct_urban	1190539.63	72.1276889	0	100.0000000	100.0000000
pct_rural	460060.37	27.8723111	0	100.0000000	0
avg_hhd_sz	45976.00	2.7854114	1.5400000	4.8500000	2.7600000
pop_sq_mile	33597928.48	2035.50	0.4630068	35142.81	1114.43
Pct_Black	71447.13	4.3285551	0	99.3517018	1.0826640
Pct_Hisp	85463.50	5.1777235	0	91.1058993	2.1297037
Pct_nonHisp_wht	1407639.03	85.2804452	0.3291880	100.0000000	91.9733285
Pct_Asian	57250.60	3.4684718	0	81.1892510	1.2513360
pct_nhpi	1666.05	0.1009357	0	46.3976288	0.0183520

cluster_new=Advd homeowner

The MEANS Procedure

Variable	Sum	Mean	Minimum	Maximum	Median
pct_api	58916.64	3.5694075	0	83.3047456	1.3002314
pct_AIAN	6633.89	0.4019075	0	51.7990496	0.2206061
pct_LIHH_span	8617.33	0.5221044	0	28.2133090	0
Pct_sing_units	1492183.81	90.4025085	61.3829357	100.0000000	91.0390826
Pct_10_units	34571.27	2.0944668	0	33.5805799	0.3891051
pct_mob_home	59669.58	3.6150235	0	38.2066277	0.5088134
pct_gq_noninst	7377.34	0.4469492	0	49.7133346	0
Pct_65plus	205107.35	12.4262300	0.6198347	88.4636413	11.5994290
pct_pop_lt18	440753.48	26.7026223	0	52.5465230	26.6059516