# A MULTIVARIATE ANALYSIS OF THE CENSUS OMISSION OF HISPANICS AND NON-HISPANIC WHITES, BLACKS, ASIANS AND AMERICAN INDIANS: EVIDENCE FROM SMALL AREA ETHNOGRAPHIC STUDIES

Manuel de la Puente, Bureau of the Census, CSMR, WPB 1, Room 433, Washington, D.C. 20233

KEY WORDS: Undercount, census errors, ethnographic research

## I. Introduction

Until very recently, relatively little information was available concerning the differential undercount of minority groups, especially Hispanics. What is known about the differential net undercount of racial and ethnic minorities has been obtained through demographic analysis (Robinson 1991; Robinson, et al. 1991 and Robinson 1988), post enumeration surveys (Hogan 1991) and, most recently, a handful of studies based on data from Census Bureau sponsored small area ethnographic studies (de la Puente 1993).

This paper is based on data from these small area ethnographic studies. This effort is known as the Ethnographic Evaluation of the Behavioral Causes of Census Undercount (hereinafter referred to as the Ethnographic Evaluation), one of the projects in the Census Bureau's Research, Evaluation and Experimental Programs for the 1990 Census. Data for the Ethnographic Evaluation were collected in 29 sample areas. Twenty eight of these sample areas were located in the continental U.S. and one was located in Puerto Rico. The sample areas were selected because they were difficult to enumerate and populated with historically undercounted minorities -- Blacks, Hispanics, Asians and American Indians.

In this paper I focus on census omissions and not net undercount calculated using the dual system estimate. My analysis is limited to individuals who were Census Day residents of the ethnographic sample areas and were missed or correctly enumerated by the 1990 census. I did not consider erroneous enumerations, a component of the net undercount.

Before I discuss my findings I provide a brief discussion of the methodology and data collection effort of the Ethnographic Evaluation. My focus is on three key aspects of this effort: sample area selection, the Alternative Enumeration, and field resolution.

## II. Background[1]

Sample area selection was driven by a number of factors including availability of qualified ethnographers. Sample areas contained about 100 housing units in one or more census blocks. Before fieldwork began each sample area was expressed in census block geography. According to the Alternative Enumeration, in all 29 sample areas combined, there are a total of 110 census blocks, 3,367 housing units and 8,718 individuals. These sample areas were purposively selected and therefore do not constitute a statistical sample. Thus, the findings reported in this paper cannot be generalized, in a statistical sense, to localities outside the sample areas.

Experienced ethnographers conducted field research for the Census Bureau under Joint Statistical Agreements. They used unobtrusive ethnographic methods, including ,participant observation, direct observation and ethnographic interviews to conduct an Alternative Enumeration (AE) of each ethnographic sample area. This entailed listing every housing unit located within the sample area, drawing a sketch map and collecting information comparable to the census "short form" (e.g., name, address, relationship to "P1" or the first person listed on the census form, sex, race, Hispanic origin, marital status and age) on every site resident. The AE phase of the project lasted approximately six weeks and had to begin within three months of Census Day (April 1, 1990).

Using census geographical codes, Census Bureau processing office clerks pulled the census forms for every housing unit in all 29 sample areas. Census questionnaires were pulled from July 1990 through October 1990. Census data were then keyed onto electronic files directly from these census forms. Census data were matched to data collected by the Alternative Enumeration (AE) for each of the sample areas.

The matching was conducted using a computer program and assisted by clerical review.[2]

The result of the matching process was a listing of matched and unmatched census and AE records for each sample area. This listing was sent to the ethnographers to use in their follow up fieldwork. In order to define the Census Day population in each of the sample areas and determine how many individuals were enumerated by the census and how many were missed or erroneously enumerated by the census, the follow up fieldwork phase of the project required the ethnographers to rule, on the basis of their intimate knowledge of the sample area and its population, on whether or not person records that matched the census with the AE were correctly or incorrectly matched. Ethnographers were also asked to explain non-matched person records, that is, why some individuals were not enumerated in the census or in the Alternative Enumeration.

**III. Summary of Findings: Cross Tabulations**

In general, my findings confirm what other studies have reported regarding the differential coverage among male, females and Blacks[3] and support the information provided by the ethnographers in their sample area reports to the Census Bureau.[4] I found that differential census omission is not only gender based but also based on race (specifically Blacks), Hispanic origin, marital status, age and relationship to "P1", the first person listed on the census form.

A typical individual omitted from the census count is likely to be male, never married as opposed to married or single (i.e., married at one time but currently widowed or divorced), between 19 and 44 years old, unrelated (e.g., border or roommate) to the first person listed on the census form in whose name the housing unit is owned or leased, Black or from a racial group other than White, that is, American Indian or Asian and Puerto Rican other Hispanic (e.g., Salvadoran or Nicaraguan) other than Mexican.

More specifically, I found that a higher proportion of Black and "other" race males and females, compared to their White counterparts, were not censused. However, overall, more males than females, regardless of race, were not counted by the census. Among Hispanics, I found that a higher proportion of Puerto Ricans and "other" Hispanics (e.g., Salvadoran, Dominicans and Guatemalans) than Mexicans were omitted from the census. Again, as with race, a relatively higher proportion of males than females, regardless of Hispanic origin, were missed by the census.

Ethnographic information collected from all 29 sample areas disclosed that reasons for within household omission include complex household structure and fear of government and of non-community members on the part of sample area residents. Complex households were found in sample areas with a sizeable concentration of recent immigrants, Blacks and American Indians. The reason for the occurrence of complex households in these sample areas included economic need, the conditions encountered by immigrants of any national origin and culturally based definition of "household" and "family" that often runs counter to the Census Bureaus' definition of household. Fear of government and of non-community residents was also a contributing factor to within household omission.

I found that within household omissions were is more likely to occur among males than females and significantly more likely to happen among single or never married individuals than those who are married. Additionally, "other" relatives or non-relatives are more likely than relatives to be within household omissions. Compared to Whites, Blacks and those in the "other" category, American Indians and Asians are more likely to be missed by the census in a partially enumerated household. Finally, within household omission is more common among Mexicans and other Hispanics than among Puerto Ricans and non-Hispanics.

**IV. Summary of Findings: Logistic Regression Analysis**

I ran a series of logistic regressions using the dichotomous response variable -- omitted from the census and correctly censused -- in order to examine the direct, combined and relative effects of demographic variables on census omission and correct census enumeration. I tested several logistic regression models and

found that the model that best fitted the sample data included the direct effects of gender, age, marital status, relationship to the household head, race and Hispanic origin as well as the interaction of gender with age, marital status, relationship race and Hispanic origin. However, these demographic characteristics, and their interactions, did not fit the data very well. Models tested using these variables had relatively high likelihood ratio chi-square statistic relative to the degrees of freedom in the model thus indicating unexplained variability. I had to include sample area effects into these model in order to fit the data. Indicating that, for these data, sample area is a necessary and important component of census omission and census enumeration.

The best and final model included the demographic variables just mentioned, and their interaction, as well as the direct effect of the sample areas. In this final model the chi-square statistic for lack of fit is 1345 with 1295 degrees of freedom. Given the relatively small sample sizes, the final model fits the data quite well. Below I discuss the results of this model.

Cross tabulations from the ethnographic sample areas as well as results from demographic analysis and post enumeration surveys show that, in general, a higher proportion of males than females are missed by the census (For example see, Fay et al. 1988 and Robinson 1991). My findings indicate that gender has no significant main effect on census omissions after controlling for all the variables in the model. Similarly age (with the exception of those age 45 or older) has no significant primary effect. The data show that those age 45 and over are significantly less likely to be missed by the census. The interaction of both gender and age has a combined or joint effect on census omission net of the direct effect of gender, age, marital status, relationship to the household head and the main effect of sample areas. More specifically, females aged 0 to 18 have significantly lower odds of being omitted from the census than their male counterparts. Females 45 years old or over also have lower odds of being missed by the census compared to males in the same age group. However, there is no statistically significant interaction of sex and age for those aged 19 to 44.

Additionally, I found that marital status, relationship to the first person listed on the census form, race and Hispanic origin have significant main effects on census omission. However, with the exception of "other relatives", the interaction of these variables with gender showed no significant combined effect on census omission. In other words, the effect of these factors on census omission is direct and independent of gender. Ethnographic observations from the sample areas suggests that single men are more likely than single women to be omitted from the census and that Black men and Hispanic men are also more likely than their female counterparts to be left off the census count. The former occurs because of economic circumstances and residential mobility and the latter because many in the sample areas were recent immigrants and in this country illegally (de la Puente 1993).

The main effect of marital status, relationship to the first person listed on the census form, race and Hispanic origin, are, with few exceptions, in line with our field observations and prior research. In these sample areas, individuals who are married or were at one time married but are currently single have lower odds of being omitted from the census than those who have never married. Those related to the first person listed on the census form have lower odds of being overlooked by the census than those who are marginally related, or not related at all, to the first person listed on the census form in whose name the housing unit is owned or leased. In these 29 sample areas, Whites and Blacks have lower odds of being omitted from the census than those in the "other" race category. This finding is unexpected regarding Blacks but anticipated with respect to "other" race given the fact that most in this racial category are Hispanics and the fact that the model shows that, across all 29 sample areas, Mexicans and other Hispanics have higher odds than non-Hispanics of being excluded from the census. Although the finding concerning Blacks is unexpected, given the information available from ethnographic field research and other research, it indicates that, in the ethnographic sample areas, the omission of

Blacks from the census is probably more complex and thus not evident given the constraints of the model.

As I mentioned earlier, the inclusion of sample areas in the model was key in fitting the model. When sample areas were included in the model the likelihood ratio chi-square was substantially reduced relative to the degrees of freedom. This suggests that sample areas add to the model's explanatory power above and beyond the contribution of demographic variables and their interactions. This finding is well supported by the independent observations of highly qualified ethnographers in all sample areas.

The ethnographers' coverage reports document specific sample area features that lead to census omissions and other erroneous enumerations that I was only crudely able to include in the model as sample area main effects. For example, crime, specifically drug dealing and use and the violence associated with these activities, were observed in a number of sample areas and declared as major obstacles to census enumeration by the ethnographers (de la Puente 1993). Irregular housing also presented problems for the census in a number of sample areas. In fact, we estimate that across all 29 sample areas as much as 40 percent of persons who should have been enumerated by the census were not because the housing unit was missed or erroneously enumerated by the census (Brownrigg 1991). These and other sample area features such as lack of affordable housing and the local economic condition are crudely represented in the model through sample area effects.

I found that about one third of the sample areas had significant main effects on census omission. Of these about half were associated with high odds of census omission. It should be kept in mind that variability across sample areas can be due, in part, to variation in the quality of the Alternative Enumerations across the 29 sample areas. This could also account for the sample area effects noted in the model.

## V. Conclusion

In general, the demographic profile of those omitted from the census across all 29 sample areas reflect the results reported by demographic analysis and post enumeration surveys. However, a systematic approach is needed to validate the findings from the Ethnographic Evaluation using statistically valid samples such as the sample of the 1990 Post Enumeration Survey (PES).

For example, patterns of census omission (and other erroneous enumerations) detected in the sample areas and validated by the 1990 PES can be further investigated using qualitative information collected by the ethnographers. For instance, if the omission of non-relatives within census households is confirmed by 1990 PES data then qualitative information in the coverage reports and behavioral information recorded by the ethnographers concerning circumstances under which within household omissions occur (e.g., concealment of information by sample area residents, shortage of affordable housing and disjunction between the Census Bureau's definition of household and what constitutes a household according to sample area residents) can be used to develop new questions for the census form, outreach messages and census enumeration procedures for the year 2000 census.

The findings presented in this paper are limited to the demographic characteristics of sample area residents and does not include data from systematic observations of the neighborhood, households and selected individuals. With respect to the neighborhood ethnographers recorded information concerning crime such as gang violence and drug use and economic conditions. Concerning household the ethnographers collected data on home language and literacy, the presence of immigrants and generations present in the household. Lastly, with respect to selected individuals, that is immigrants, ambiguous household members, and those who do not speak English well, the ethnographers collected information concerning country of birth, time of immigration, extent of residential mobility, and so on. These data can provide further insight into why people are missed or erroneously counted by the census.

Recently a working group called the Ethnographic Data Analysis Working Group was formed at the U.S. Census Bureau, Center for

Survey Methods Research to analyze these data and conduct comparative analyses using data from the Ethnographic Evaluation and the 1990 Post Enumeration Survey.

## References


Brownrigg, Leslie (1991), "Irregular Housing and Housing Counts in American Minority Neighborhoods: Preliminary Findings From the Ethnographic Evaluation," Unpublished paper, U.S. Census Bureau, Center for Survey Methods Research.

Brownrigg, Leslie and Manuel de la Puente (1993), "Alternative Enumeration Methods and Results: Resolution and Resolved Populations by Site." Preliminary Research and Evaluation Memorandum (PREM) No. 219.

de la Puente, Manuel (1993), "Why Are People Missed or Erroneously Included by the Census: A Summary of Findings From Ethnographic Coverage Reports." Paper presented at the U.S. Census Bureau's Research Conference on Undercounted Ethnic Populations, Richmond, VA, May 1993.

de la Puente, Manuel (1991), "In Search of the Causes of the Differential Census Undercount of Racial and Ethnic Minorities: Overview of Ethnographic Studies of Census Undercount." Paper presented at the 86th annual meeting of the American Sociological Association, Cincinnati, Ohio, August 1991.

Fay, Robert et al. (1988), The Coverage of Population in the 1980 Census. Bureau of the Census Evaluation and Research Reports PHC80-E4. Washington, D.C.: U.S. Government Printing Office.

Hogan, Howard (1991), "The 1990 Post-Enumeration Survey: Operations and Results." Paper presented at the annual meeting of the American Statistical Association, Atlanta, Georgia, August 1991.

Robinson, Gregory (1991), "Results of the Evaluation of Coverage in the 1990 Census Based on Demographic Analysis." Memorandum for Paula Schneider, Chief, Population Division, U.S. Census Bureau, June 4, 1991.

Robinson, Gregory et al. (1991), "Estimating Coverage of the 1990 United States Census: Demographic Analysis." Paper presented at the annual meeting of the American Statistical Association in Atlanta, Georgia, August 1991.

Robinson, Gregory (1988), "Perspectives on the Completeness of Coverage of the Population of the United States Decennia Census." Paper presented at the annual meeting of the Population Association of America, New Orleans, Louisiana, April 1988.

Slaven, Bradley (1990), "The Set Theory Matching System." Paper presented at the University of Georgia Advanced Computing and Information Technologies for the Social Sciences Conference, Athens, Georgia, April 1991.


## NOTES

1. The views expressed are attributable to the author and do not necessarily express those of the U.S. Bureau of the Census.

2. For more detailed background information on the ethnographic evaluation see de la Puente (1991) and Brownrigg and de la Puente (1992).

3. The matching was conducted using a computer matching program developed specifically for the project. For more information see Slaven (1991).

4. For example see Robinson (1991); Robinson et al. (1991) and Fay et al. (1988).

5. All 29 coverage reports are available from the U.S. Census Bureau, Center for Survey Methods Research, Washington, D.C. For a summary of findings presented in the coverage reports see de la Puente (1993).

TABLE 1
DEMOGRAPHIC CHARACTERISTICS BY ENUMERATION STATUS
(SAMPLE SIZE IN PARENTHESIS)

| DEMOGRAPHIC CHARACTERISTICS | ENUMERATION STATUS MISSED | CENSUSED |
|---|---|---|
| SEX | | |
| MALE | 18.5% (692) | 81.5% (3047) |
| FEMALE | 14.1 (525) | 85.9 (3191) |
| MATITAL STATUS | | |
| MARRIED | 13.2 (294) | 86.8 (1936) |
| NOT MARRIED | 20.2 (725) | 79.8 (2871) |
| SINGLE | 8.9 (129) | 91.1 (1294) |
| AGE | | |
| 0-6 | 17.6 (174) | 82.4 (815) |
| 7-14 | 14.8 (156) | 85.2 (896) |
| 15-18 | 13.8 (76) | 86.2 (475) |
| 19-29 | 20.9 (314) | 79.1 (1186) |
| 30-44 | 16.6 (284) | 83.4 (1427) |
| 45-64 | 14.3 (154) | 85.7 (919) |
| 65 & OVER | 10.2 (59) | 89.8 (520) |
| RELATIONSHIP | | |
| RELATIVE | 14.2 (939) | 85.8 (5671) |
| OTHER RELATIVE | 27.0 (85) | 73.0 (230) |
| NON-RELATIVE | 36.6 (186) | 63.4 (322) |
| RACE | | |
| NON-HISPANIC WHITE | 10.2 (106) | 89.8 (932) |
| NON-HISPANIC BLACK | 19.1 (346) | 80.9 (1462) |
| NON-HISPANIC AMERICAN INDIAN | 11.9 (98) | 88.1 (727) |
| NON-HISPANIC ASIAN PACIFIC ISLANDER | 11.8 (115) | 88.2 (727) |
| NON-HISPANIC OTHER RACE | 23.6 (55) | 76.4 (178) |
| HISPANIC ORIGIN | | |
| MEXICAN | 14.9 (216) | 85.1 (1237) |
| PUERTO RICAN | 25.8 (104) | 74.2 (299) |
| OTHER HISPANIC | 24.3 (174) | 75.7 (541) |
| NON-HISPANIC | 14.8 (721) | 85.2 (4156) |

| | |
|---|---|
| SEX | $X^2=26.1$; df=1; <.05 |
| MARITAL STATUS | $X^2=114.1$; df=2; <.05 |
| AGE | $X^2=47.9$; df=6; <.05 |
| RELATIONSHIP | $X^2=201.5$; df=2; <.05 |
| RACE | $X^2=71.1$; df=4; <.05 |
| HISPANIC ORIGIN | $X^2=70.1$; df=3; <.05 |

TABLE 2
DEMOGRAPHIC CHARACTERISTICS BY WHOLE AND WITHIN HOUSEHOLD OMISSION
(SAMPLE SIZE IN PARENTHESIS)

| DEMOGRAPHIC CHARACTERISTICS | OMISSION WHOLE | WITHIN |
|---|---|---|
| SEX | | |
| MALE | 61.9% (422) | 38.1% (260) |
| FEMALE | 69.0 (357) | 31.0 (160) |
| MATITAL STATUS | | |
| MARRIED | 74.4 (215) | 25.6 (74) |
| NOT MARRIED | 63.6 (454) | 36.4 (260) |
| SINGLE | 63.2 (79) | 36.8 (46) |
| AGE | | |
| 0-6 | 57.3 (98) | 42.7 (73) |
| 7-14 | 67.1 (102) | 32.9 (50) |
| 15-18 | 64.9 (48) | 35.1 (26) |
| 19-29 | 57.4 (178) | 42.6 (132) |
| 30-44 | 66.5 (187) | 33.4 (94) |
| 45-64 | 80.9 (123) | 19.1 (29) |
| 65 & OVER | 72.9 (43) | 27.1 (16) |
| RELATIONSHIP | | |
| RELATIVE | 72.5 (671) | 27.5 (254) |
| OTHER RELATIVE | 43.5 (37) | 56.5 (49) |
| NON-RELATIVE | 37.4 (68) | 62.6 (114) |
| RACE | | |
| NON-HISPANIC WHITE | 70.9 (73) | 29.1 (30) |
| NON-HISPANIC BLACK | 78.0 (269) | 22.0 (76) |
| NON-HISPANIC AMERICAN INDIAN | 59.1 (52) | 40.9 (36) |
| NON-HISPANIC ASIAN PACIFIC ISLANDER | 53.1 (60) | 46.9 (53) |
| NON-HISPANIC OTHER RACE | 69.1 (38) | 30.9 (17) |
| HISPANIC ORIGIN | | |
| MEXICAN | 51.2 (110) | 48.8 (105) |
| PUERTO RICAN | 70.2 (73) | 29.8 (31) |
| OTHER HISPANIC | 59.0 (102) | 41.0 (71) |
| NON-HISPANIC | 69.8 (492) | 30.2 (213) |

| | |
|---|---|
| SEX | $X^2=6.6$; df=1; <.05 |
| MARITAL STATUS | $X^2=11.46$; df=2; <.05 |
| AGE | $X^2=31.4$; df=6; NS |
| RELATIONSHIP | $X^2=101.5$; df=2; <.05 |
| RACE | $X^2=30.4$; df=4; <.05 |
| HISPANIC ORIGIN | $X^2=29.2$; df=3; <.05 |

Table 3
MULTIVARIATE LOGISTIC REGRESSION ANALYSIS OF CENSUS OMISSION

| VARIABLES | OMISSION PAR | $X^2$ | PROB |
|---|---|---|---|
| A) SEX (df=1) | | | |
| MALE | .0997 | 1.70 | NS |
| FEMALE | -.0997 | 1.70 | NS |
| B) AGE (df=2) | | | |
| 0-18 | .0760 | 1.33 | NS |
| 19-44 | .0980 | 3.51 | NS |
| 45-& OVER | -.1740 | 5.70 | .0162 |
| C) MARITAL STATUS (df=2) | | | |
| MARRIED | -.0398 | 0.37 | NS |
| NEVER MARRIED | .1734 | 6.54 | .0105 |
| SINGLE | -.1335 | 2.60 | NS |
| D) RACE (df=2) | | | |
| WHITE | -.1635 | 4.52 | .0335 |
| BLACK | -.2113 | 4.37 | .0273 |
| OTHER RACE | .3748 | 27.20 | .0000 |
| E) HISPANIC ORIGIN (df=2) | | | |
| MEXICAN | .3092 | 9.02 | .0027 |
| OTHER HISPANIC | .0470 | 0.28 | NS |
| NON-HISPANIC | -.3562 | 13.04 | .0003 |
| F) RELATIONSHIP (df=2) | | | |
| RELATIVE | -.4819 | 46.73 | .0000 |
| OTHER RELATIVE | .1634 | 2.47 | NS |
| NON-RELATIVE | -.3185 | 10.70 | .0010 |
| G) SEX/AGE INTERACTION (df=2) | | | |
| MALE 0-18 | -.1619 | 6.45 | .0111 |
| MALE 19-44 | .0119 | 0.05 | NS |
| MALE 45 & OVER | -.1500 | 4.60 | .0320 |
| H) SEX/MARITAL STATUS INTERACTION (df=2) | | | |
| MALE MARRIED | -.0313 | 0.24 | NS |
| MALE NEVER MARRIED | .0750 | 1.48 | NS |
| MALE SINGLE | .0438 | 0.34 | NS |
| I) SEX/RELATIONSHIP INTERACTION (df=2) | | | |
| MALE RELATIVE | .0402 | 0.33 | NS |
| MALE OTHER RELATIVE | -.2195 | 4.51 | .0337 |
| MALE NON-RELATIVE | -.1793 | 3.44 | NS |
| J) SEX/RACE INTERACTION (df=2) | | | |
| MALE WHITE | .0041 | 0.00 | NS |
| MALE BLACK | .0332 | 0.29 | NS |
| MALE OTHER RACE | .0374 | 0.57 | NS |
| K) SEX/HISPANIC ORIGIN INTERACTION (df=2) | | | |
| MALE MEXICAN | .0138 | 0.04 | NS |
| MALE OTHER HISPANIC | .0794 | 1.60 | NS |
| MALE NON-HISPANIC | .0932 | 2.83 | NS |

NOTE: ALSO INCLUDED IN THE EQUATION AS CONTROLS, BUT NOT SHOWN, ARE SAMPLE AREA EFFECTS.