BUREAU DOCUMENTATION

1995 Census Test -- Sampling Methods

1. Introduction

A census, like any statistical activity, is subject to some error. In the past, population counts from a census were primarily subject to non-sampling errors (respondent error, failure to count everyone, processing error, etc.) and not sampling error since sampling was not used in producing population totals.

For Census 2000, the Census Bureau plans to introduce many improvements to significantly reduce the non-sampling error of past censuses, particularly, the error from failure to count everyone. The process for Census 2000 will incorporate a combination of improved counting procedures and the use of statistical estimation, meaning the census population counts will have significantly less non-sampling error than in previous censuses, but there will now be some measurable sampling error associated with population totals. Major components of the Census 2000 process were tried in the 1995 test censuses held in Oakland, CA, Paterson, NJ, and NW Louisiana which means the tabulations are subject to some sampling error.

Section 2 below describes how to calculate the sampling error for the 1995 Census Test data. Section 3 describes the sampling and estimation methods used in the 1995 Census Test.

2. Calculation of Sampling Errors

For Estimated Totals - Standard errors are used as a measure of the sampling error. Standard errors for selected population totals are given in Table A. Standard errors are provided for each site because sampling methodology was applied differently to each site as part of a test to identify the best methodology for Census 2000.

Example of how to use Table A: For Block 401 within Tract 407900 in Oakland, CA, the estimated total population is 300. To calculate the standard error for this estimate, find the estimate 300 in the left column of Table A. Follow across the row and find the associated standard error for Oakland. This value is 26. So, the error associated with the estimate is 26.

To understand how the standard error effects the estimates, calculate a confidence interval. This will give a range around the estimate that will contain the "true" population total, ninety percent of the time.

Example of how to calculate Confidence Interval: Using the example above, a 90 percent confidence interval around the estimated population total is 300 plus or minus 1.645 * 26 or (257 to 343). So, with 90% probability, the "true" number of people 18+ is between 257 and 343.

For population values not given in the table, standard errors may be approximated. For example, for Block 401 within Tract 407900 in Oakland, the estimated total population of Asian and Pacific Islanders is 75. The standard error for a population estimate of 75 is not listed in Table A, however, since 75 is 1/2 of the distance between population values 70 and 80, the standard error is about 1/2 of the distance between the corresponding standard errors, 12 and 13, which is about 12.5. Alternatively, standard errors for the estimated number of people for a characteristic of interest may be calculated using the formula provided at the bottom of Table A.

For Estimated Percentages - The calculations are similar to the above. Standard errors for a percentage can be found in Table C for Oakland, CA; in Table D for Paterson, NJ; and in Table E for Northwest Louisiana.

Example of how to use Table C: For Block 401 within Tract 407900 in Oakland, the estimated percentage of Asian and Pacific Islanders is 25% with an estimated base population equal to 300. To calculate the standard error of the estimated percentage, use Table C and find the estimated base population (300 for this example) in the left column. Follow the row across to the column containing the estimated percentage, 25%, to find the standard error equal to 3.59%. So, for this example, the estimated 25% has a standard error equal to 3.59%.

Example of how to calculate Confidence Interval: Using the example above, a 90 percent confidence interval for the estimated percentage is about equal to 25% plus or minus 1.645 * 3.59% or (19% to 31%). So, with 90% probability, the "true" percentage of people that are Asian or Pacific Islander is between 19% and 31%.

For percentages not given in the table, standard errors may be approximated. Alternatively, standard errors for the estimated percentage of people that have a certain characteristic may be calculated by using the formula given at the bottom of each table.

3. 1995 Census Test -- Sampling Methods

This census test used sampling in two new ways. First, after repeated attempts to enumerate all households either by mail, telephone or personal visit and as time and resource limits approached; a sample of all the addresses that had not responded was selected. A census worker visited and interviewed each of the sample addresses and obtained a list of all the residents in the occupied housing units. This use of sampling is called "Sampling for Nonresponse". The results of this sample were used to estimate the population in nonresponding household.

A second new use of sampling was carried out as a quality check on the census results. A small sample of all addresses in the test sites was chosen and census interviewers visited each address, enumerated the persons living there and determined if they had been previously counted in the census test. The results of this check (called "Integrated Coverage Measurement") were used to estimate the number of people "missed" or "double-counted" in the actual census. These results were incorporated in the population estimates for the test site.

Sampling methodology for nonresponse and Integrated Coverage Measurement was applied to each test site. Both these sampling techniques have associated, measurable "error." Therefore, the final population totals from the 1995 Census Test are estimates subject to some range of uncertainty, measured by the standard errors. For the 1995 Census Test, the statistical relationship between estimated totals and their associated standard error estimates was modeled for each site. Because the standard error behavior was not identical for all estimates within a site, the standard errors we computed from the variance model parameters provide an approximation of the standard error for any specific estimate.

For technical assistance or more details on the sample design and estimation procedures, please contact Alfredo Navarro at (301) 457 - 1962.

Table A. Standard Errors for Population Totals

Total Population	Oakland	Paterson	NW Louisiana
10	5	8	5
20	6	11	7
30	8	13	8
40	9	15	10
50	10	17	11
60	11	19	12
70	12	20	13
80	13	22	14
90	14	23	14
100	14	24	15
200	21	35	22
300	26	42	27
400	30	49	31
500	33	55	35
600	37	61	38
700	40	66	41
800	43	70	44
900	46	75	47
1000	49	79	50
2000	75	114	73
3000	97	144	93
4000	118	169	111
5000	139	194	128
6000	159	217	144
7000	179	239	160
8000	199	260	175
9000	219	281	190
10000	238	301	205
20000	431	494	351
30000	623	679	492
40000	813	861	633
50000	1004	1041	773
60000	1195	1220	912
70000	1385	1399	1052
80000	1575	1578	1191
90000	1766	1756	1330
100000	1956	1934	1469
200000	3859	3713	2860
300000	5762	Not Applicable	Not Applicable
400000	7665	Not Applicable	Not Applicable

Formula: s.e.(x) = sqrt{(a)*(x**2) + (b)*(x)}

where, x = estimated number of people, and a and b are parameters given in Table B.

Table B. Variance Model Parameters for the Calculation of Standard Errors

	Parameters	
Census Sites	a	b
Oakland	.000362	2.061328
Paterson	.000315	5.922080
NW Louisiana	.000193	2.292658