2010 Census Coverage Measurement : The Hunt for the Magic Variables

Eric Schindler
U.S. Census Bureau, Washington DC 20233

Abstract

The Census 2000 Accuracy and Coverage Evaluation (A.C.E.) employed a post-stratified dual system estimation design. The key post-stratification variables were race, Hispanic origin, age, sex, and tenure, familiarly known as ROAST. Additional variables used in A.C.E. were Census region, type of enumeration area, tract mail return rate, family type, return type / return date / proxy status, state, and tract percent minority. Current research for the 2010 Census Coverage Measurement (CCM) program is investigating logistic regression based dual system estimators with the ROAST variables and additional variables. This paper examines the differences several of these additional variables make in the dual system estimates at different levels of aggregation such as for small areas or for race / origin groups once ROAST is included in the post-stratification model.

Key Words : post-stratification, dual system estimation

1. Background

Since 1970 the U.S. Census Bureau has measured coverage in the decennial census using a post-stratified dual system estimator (DSE). The post-stratification variables included race, Hispanic origin, age, sex, and tenure, familiarly known as ROAST, and additional variables for the larger ROAST groups. For the 1990 Post Enumeration Survey (PES) and the 2000 Accuracy and Coverage Evaluation (A.C.E.), collapsing to maintain minimum sample sizes resulted in many post-strata for non-Hispanic whites, fewer post-strata for non-Hispanic Blacks and Hispanics, and only the basic ROAST design for Native Hawaiians and Pacific Islanders (NHPI), Asians, American Indians / Alaska Natives living on reservations (AIR), and American Indians / Alaska Natives not on reservations (AInR).

For the PES and the A.C.E., the basic design involved the selection of a sample of block clusters spread across the country. Census reports without at least a minimum amount of reported data were classified as non data-defined. The remaining data-defined census reports (DD) in these block clusters were combined to form the Enumeration or E sample. A second independent collection in these block clusters was designated as the Population or P sample. The E sample was matched to the P sample and E-sample persons (E) found in the P sample were generally classified as correct (CE) enumerations. The P sample was matched to the data defined census records in the block clusters and most matched P-sample records (P) were classified as residents as of Census Day and as matched (M). In some cases matches to a slightly larger search area than the block cluster were allowed. A field follow-up operation was performed to resolve unmatched and possibly matched cases. For some cases it was necessary to impute a probability of whether an E-sample person was a correct or erroneous enumeration or whether a P-sample person report corresponded with a Census Day resident or whether a P-sample resident was a match or a non-match in the search area. The E-sample erroneous enumerations were intended to estimate the number of records in the census which should not have been collected because they did not correspond to a Census Day resident in the search area while the P-sample non-matches were intended to estimate the number of persons who were Census Day residents but were missed by the census in the search area.

The basic form of the DSE for any post-stratum is:

$$DSE = DD \frac{r_{CE}}{r_{Match}} = DD \frac{CE}{E} \frac{P}{M}$$

where the term $DD \frac{CE}{E}$ is the weighted estimate of the number of correctly enumerated persons for whom it was possible to attempt a match. P excludes any Census Day non-residents identified in the P sample, r_{CE} is called the correct enumeration rate, and r_{Match} is called the match rate. Slightly more complicated formulations for P and M were used for the A.C.E. in order to handle movers. Under the appropriate assumptions, the DSE estimates the true population of the A.C.E. universe which excluded people in group quarters and remote areas of Alaska.

A coverage correction factor, or CCF, was calculated for each post-stratum by dividing the DSE by the census count. Small area or population estimates were calculated by adding the products of the census counts in each post-stratum in the subpopulation and the post-stratum CCFs. A.C.E. variances were calculated by a stratified Jackknife procedure dropping one block cluster at a time.

Analysis of the A.C.E. estimates before the scheduled

———————————

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical and methodological issues are those of the author and not necessarily those of the U.S. Census Bureau.

release in March 2001 suggested serious problems, especially when compared to the Census Bureau's demographic analysis (DA) estimates. Later investigation indicated that there were several million duplicated people in the census and that several million more weighted persons should have been classified as erroneously enumerated. P- and E-sample records with possible duplicates outside the search area and an evaluation sample containing additional information for about 10 percent of the 700,000 E- and P-sample records were sent to the Jeffersonville, Indiana processing center for recoding. This effort, called A.C.E. Revision II, used different post-stratification designs for the E and P samples, separated out the possible duplicate records, and introduced double sampling adjustments for the nonduplicated records to adjust the original correct enumeration and match rates by the changes made to these rates by the recoding of the subsample. It resulted in a much more complex formula but still of the basic form:

$$DSE_{ACErevII} = DD \frac{r_{CE,\ ACErevII}}{r_{Match,\ ACErevII}}$$

Because only a 10 percent sample was available for the double sampling adjustments for measurement error, the estimates are more volatile and the variances were approximately doubled from those observed for A.C.E. and expected for the CCM program in 2010. Variances are calculated by a simple Jackknife which yields similar results to the stratified Jackknife used in production.

U.S. Census Bureau (2004) serves as a general reference for background information for this entire paper.

2. Objective

For the 2010 census, the goals of the Census Coverage Measurement program or CCM, have evolved. In addition to the traditional emphasis on net coverage error, the Census Bureau is also interested in identifying and measuring the components of coverage error:

- the number of people erroneously enumerated in the census
- the number of people missed by the census

This added objective of CCM will permit the analysis of coverage errors in the 2010 census operations so that subsequent censuses can be improved.

Current research with the A.C.E. data will allow the Census Bureau to assess the use of logistic regression modeling to assign each census record an individual CCF. Modeling would allow the use of more variables than are possible with post-stratification resulting in improved estimates to better explain coverage variation. See Griffin, Mule, and Olson (2006).

The objective of this paper, working from a post-stratification foundation and with net coverage estimates, is to discuss the different types of variables that might be used, their advantages and disadvantages, and to show some results of adding these variables to the basic ROAST design. The ideal is to find the magic variables that explain census coverage variation and can be defined from the limited census and independent sample data and included in the logistic regression design

Estimates for both the A.C.E. and A.C.E. Revision II designs were compared and the differentials between population groups for different post-stratification designs were similar. Thus, even though the original A.C.E. estimates are not appropriate for the level of net coverage error, they are appropriate for comparing differences in the estimates. Considering that their variance structure is similar to that expected for 2010, they are actually better than A.C.E. Revision II for evaluating the statistical significance of differences. This paper presents the effects of including different variables in the post-stratification design using the March 2001 A.C.E. data.

3. Variables

This paper assumes that the ROAST variables will continue to be important covariates because (1) they have proven useful in measuring coverage - Blacks and Hispanics are harder to count than non-Hispanic Whites; younger males are harder to count than older females; renters are harder to count than owners and (2) there is public interest in direct estimates for these subgroups.

Additional variables which have been used over the past several censuses or which are being considered for use in 2010 fall into several categories:

- Broad geographic variables: Examples are the four Census Regions or the Metropolitan Statistical Area or the Type of Enumeration Area (MSA/TEA) variable used in 2000: large MSA mailout/mailback (mo/mb) areas, medium MSA mo/mb areas, other mo/mb areas, all non mo/mb areas.
- Tract-level variables: These variables identify tracts (there are 60,000 total in the U.S.) which may be difficult to enumerate accurately. Examples are the tract-level percentage minority or mail return rate. These variables are correlated so it may not make much difference which one is used.
- Family variables: Married householders (HHer) and their spouses are easier to count correctly than a boarder renting a room. One measurement problem is that the relationships may change between the census and the coverage sample if a different resident is identified as the HHer.
- Census operational variables: The A.C.E. Revision II used a five-category variable for the E sample with early mail return form, late mail return form, early

nonresponse follow-up form, late nonresponse follow-up form, or proxy report. Although useful for explaining variation in correct enumeration rates, this variable could not be defined for nonmatched P-sample housing units. A.C.E. Revision II addressed this problem by the use of different post-stratification designs for the E and P samples. One unintended consequence was that A.C.E. Revision II estimated large census overcounts for a number of small places.

4. Empirical Design

This paper compares the estimates for sixteen different post-stratification models for the A.C.E. estimates:

- 1 cell : This simplest design shows the effect of basic dual system estimation without post-stratification.
- 7 cells : one cell for each of the race/Hispanic origin groups cited above
- 2 cells : tenure : one cell for owners, one for renters.
- 8 cells : one cell for each of eight age/sex categories: defined by (1) age 0-9, (2) age 10-17, (3) age 18-29 male (M), (4) age 18-29 female (F), (5) age 30-49 M, (6) age 30-49 F, (7) age 50+ M, and (8) age 50+ F.
- 14 cells : cross tenure (2) by race/Hispanic origin (7)
- 56 cells : cross race/origin (7) by age/sex (8). (Because of some variance problems the NHPI age 18 and over, age/sex cells are collapsed into adult males and adult females throughout this study.)
- 16 cells : cross tenure (2) by age/sex (8).
- ROAST : 104 cells : cross race/origin (7) by tenure (2) by age/sex (8). Before the collapse of adults for NHPI, there were 112 cells.
- M2K : This modified 2000 design approximates the design used in the A.C.E. There are 64 post-stratum groups defined within the 14 race/origin/tenure groups crossed by the 8 age/sex groups. Region, high or low tract mail return rate, and MSA/TEA are maintained for White owners and partially maintained for White renters, Blacks, and Hispanics. Collapsing of the adults occurs if there are fewer than 100 P-sample persons (nonmovers plus outmovers) in any post-stratum.
- REGION : cross ROAST and Census Region. There is no additional collapsing for Whites, Blacks, or Hispanics but considerable collapsing of the regions for the smaller groups.
- StateN : cross ROAST by STATE. With the sample sizes of the A.C.E. program the standard errors of direct state estimates are too high for general use. However, the comparison of direct and synthetic state estimates can be used to examine the general issue of synthetic error for small areas. Any ROAST by state cell with at least 100 P-sample persons is kept as a post-stratum. One additional post-stratum is defined for each ROAST cell for all the smaller states. For non-Hispanic Whites only a few cells collapse; for Blacks and Hispanics the cells in states with small minority populations collapse; for the other groups almost all cells collapse. The proportion of each state's total population not collapsed ranges from 60 percent for smaller states to 97 percent for larger states. A State variable is not feasible for a post-stratification design, but may be possible in a logistic regression design.
- StateR : cross ROAST by State. StateR collapses to the region instead of nationwide when the state sample sizes in a ROAST cell were too small.
- %BH10 : cross ROAST with a tract-level indicator of the percentage Black or Hispanic with ten values, 0-10 percent, 10-20 percent, etc. Collapsing occurs frequently for the smaller race/origin groups. Most non-Hispanic non-Blacks live in tracts that are less than 10 percent Black or Hispanic. Blacks and Hispanics are more uniformly distributed.
- %BH3 : cross ROAST with a tract-level indicator of the percentage Black or Hispanic with three values, 90-100, 50-90, or 0-50 percent Black and Hispanic for Blacks and Hispanics, and 0-10, 10-50, or 50-100 percent Black or Hispanic for the other groups.
- FAM8 : cross ROAST with a family variable which indicates the person's attachment to the household: (1) householder (HHER) living with spouse and minor children and the spouse, (2) HHer living with spouse without minor children and the spouse, (3) HHer with minor children not living with spouse, (4) HHer not living with spouse with no minor children, but with at least one other person in the household, (5) children under 30 of a HHer living with spouse, (6) children under 30 of a HHer not living with spouse, (7) all other persons, and (8) HHer living alone.
- FAM3 : cross ROAST with a family variable with three values : (1) HHer living with spouse, the spouse, and any minor children, (2) HHer not living with a spouse but living with minor children and the minor children, and (3) all other persons.

5. Results and Analysis

NOTE 1 : Two concepts of significance will be used when discussing differences between ROAST and the other designs. The first, statistical significance at the 95% level, indicates whether the difference between two estimates is greater than 1.96 times the standard error of the difference. The second, practical significance at the 95% level, indicates whether the alternate estimate falls outside the 95% confidence interval of the ROAST estimate. Highly correlated estimates have low variances of differences and can be only statistically significant; estimates with low or negative correlation can be only practically significant. In the tables differences between estimates and the corresponding ROAST estimates which are statistically significant are shown in **bold**; practically significant in *italics*; both statistically and practically significant in ***bold italics***.

NOTE 2 : The March 2001 A.C.E. estimates indicated an undercount by the census of 3 million persons; the preliminary estimates for A.C.E. Revision II indicated an overcount of 3 million persons. A broad based correction for correlation bias in A.C.E. Revision II based on DA sex ratios left an estimated overcount of about 1.3 million persons. This last estimate is considered the best estimate of the population as of April 1, 2000.

NOTE 3: This analysis views collapsing as an imputation of the correct enumeration and match rates and applies these rates to the uncollapsed data-defined rate. A low post-stratum data-defined rate may be an indication of a high correct enumeration rate (because the non-data-defined people don't have enough data to even make it into the E-sample to be erroneous enumerations) and/or a low match rate (because there are fewer data-defined census people for the P-sample people to match to). The alternative of also collapsing the data-defined rates implies that the collapsed post-strata are all the same.

5.1 Population Totals

The A.C.E. universe excluded about 8 million persons living in group quarters such as nursing homes, college dormitories, prisons, or military barracks, and about 50,000 persons living in remote areas of Alaska. The census count for the remaining population was 273,586,997. The A.C.E. production estimate was 276,848,873, within 2,000 of this study's M2K design. The difference is because children 0-17 years old were combined in the original A.C.E.

The idea of post-stratification is to form cells which have homogeneous coverage properties; that is, all persons in the cell have equal probabilities of being correctly enumerated in or of being omitted from the census. Because of the division by the match rate, a design with cells which are more homogeneous usually results in an estimate of the total population which is slightly higher than the estimates from designs with less homogeneous cells. An examination of the total population estimate for the models gives a rough idea about the effectiveness of additional post-stratification variables in improving homogeneity. Of course, other factors affect the estimates and other criteria can be used to evaluate estimates. This paper does not imply that higher estimates are automatically better estimates, only that higher estimates may be a result of more homogeneous post-strata.

Chart 1 shows national estimates for the designs. Within each box, the second line gives the estimate and the third line the difference from the ROAST estimate. Between the boxes for two models, the first line gives the additional variable in the model and the second line the change between the estimates of the two models. The one-cell DSE differs substantially from the census count, adding about 3 million people, by far the biggest difference found in this analysis. The more complicated models add up to about 500,000 to the one-cell estimates, only one sixth of the change made by the one-cell.

Adding race/origin to the one-cell model to create the 7-cell model adds about 290,000 persons to the national estimate, indicating that the 7 post-strata may be more homogeneous than a single post-stratum and that this variable may improve the estimates. Adding tenure or age/sex to create the 2-cell or 8-cell models results in smaller increases, indicating the race/origin may be the most important variable for improving homogeneity. Including two variables in the post-stratification models to form the 14-, 16-, or 56- cell models adds to the estimates again with race/origin again the strongest variable. Adding the last variable produces the ROAST estimate, almost half a million higher than the one-cell estimate.

The M2K design adds 34,000 to the ROAST estimate. The REGION estimate adds almost as much. The StateN and StateR national estimates are slightly lower.

The %BH10 and FAM8 designs resulted in national estimates less than even the one-cell estimate. Splitting the ROAST post-strata by these variables may be increasing instead of decreasing heterogeneity or may be interacting with the data-defined and correct enumeration rates in some unexpected way. In any event, these estimates and the FAM3 estimate appear to be out of line. The %BH3 estimate is about the same as the ROAST estimate and could help for lower-level estimates so there may be some useful information in a tract-level variable but only if a few classes are used.

Since it is unlikely that a design without the ROAST variables will be employed for the 2010 Census, the discussion from this point forward will focus on ROAST and the larger designs.

5.2 Post-strata

Table 1 shows the number of post-strata in each design containing ROAST and how many are significantly different, practically different, and both from the ROAST estimate. Adding variables splits post-strata and changes the estimates in the practical sense, but the standard errors of the smaller post-strata are often too high for the differences to be statistically significant. Overall about 10 percent of the post-strata are significantly different from

ROAST, 40 percent are practically different, and 2 percent are both. Thus, post-stratification below ROAST may be making some useful distinctions.

Table 1: Post-stratum Results

Design	Post-strata	Significant	Practical	Both
M2K	512	11	285	11
Region	400	87	100	16
StateR	4641	572	2210	109
StateN	4641	380	1751	53
%BH3	304	21	65	5
%BH10	976	23	432	4
Fam3	304	38	96	13
Fam8	568	65	244	32

Some other observations (not shown) are:

- The data-defined rate increases as the percentage Black or Hispanic in the tract decreases. These rates have no variance and the increases are almost monotone. The correct enumeration and match rates also trend upwards, but they bounce around. Once the three rates are combined to estimate the CCF there is no longer a clear trend. The same pattern of correlation is repeated for the family variable.
- For the state estimates, the collapsing of the smaller states gives one (StateN) or four (StateR) estimates of the correct enumeration and match rates, but the different data-defined rates result in different CCFs. For example, the StateN and StateR estimated CCFs for the states with few young Black male renters differ by as much as 0.08. The only difference is whether the collapse occurs to the national level or to the Census Region level. This is a variance/bias tradeoff between very small sample sizes with narrow regional collapsing and slightly larger sample sizes with broader national collapsing.

5.3 Demographic Groups

There is public interest in the estimates for broad demographic groups. Post-stratification variables can parse the differences between groups. Table 2 at the end of the paper shows the resulting observed coverage correction factors for the race/Hispanic origin groups and for owners and renters within each group. Although sometimes statistically significant for the %BH10, FAM8, and FAM3 designs and for two additional estimates for Asians, the differences are not practically significant.

5.4 Sex Ratios

Generally males are undercounted in the census as compared to females when compared to the DA sex ratios. This effect is called correlation bias and is caused by the fact that the systems, although operationally independent, are not functionally independent. Within a post-stratum, some males are more likely to be missed than others both in the census and in the P sample. With stronger attachment to households and fewer reasons to want to avoid enumeration, correlation bias is less of a problem for females when compared to DA.

Table 3 shows the sex ratios (number of males over number of females). The observed census sex ratios are all less than the DA target sex ratio.

The ROAST sex ratios are greater than the census sex ratios, meaning that the DSE is having some success in estimating more missed adult males or eliminating more erroneously enumerated adult females. The "effect" column shows $(SEXR_{ROAST}-SEXR_{Cen})/(SEXR_{DA}-SEXR_{Cen})$. For non-Blacks age 18-29 ROAST closes 130 percent of the gap between the census and the DA sex ratios resulting in an overcount of males relative to females. For the other non-Black cells, the differences compared to census are almost half of what would be needed to reach the DA targets. However, the effect for Blacks is small.

Post-stratification moves the sex ratios only part of the way to the DA targets so there is residual correlation bias, especially for Black adult males. With the limited range of variables available for post-stratification it does not seem possible to realize the DA target sex ratios for Blacks. DA does not have the data to separate Hispanics from non-Blacks, but Hispanics may be subject to the same correlation bias problems as Blacks.

5.5 State Estimates

In order to observe the effects of the additional variables beyond ROAST on small area estimates, synthetic A.C.E. estimates for the 51 states were calculated for the post-stratification models. The results are shown in Graph 1.

- Many of the state estimates with the REGION design and many of the Midwest estimates with the StateN and StateR designs are both statistically and practically different from the ROAST estimate. The StateN and StateR estimates are close to each other but not necessarily close to the REGION estimate, meaning that REGION captures only some of the geographic effect. The state estimates, however, are subject to much higher variances making them impractical for general use.
- There is also statistical significance between ROAST and the M2K design for eight states, several in the Midwest. All of these differences are also practically significant. It is likely that the region variable for non-Hispanic White owners included in the M2K design is making this difference.
- Almost all of the %BH10, FAM8, and FAM3 state

estimates are statistically different from the corresponding ROAST estimates but only three of the differences are also of practical significance. This is caused by the lower national estimates for these designs and the high correlation with the ROAST estimates at the state level resulting in small variances of the relatively large differences. The %BH3 design national estimate is closer to the ROAST design national estimate and no estimate is statistically or practically different.

Table 4 summarizes these results and shows the average weighted coefficients of variation (CV). The number of significant differences for the M2K and Region designs justifies the use despite the higher CVs, but the same is not true for the StateN and StateR designs.

Table 4: State Level Estimates

Model	Stat Sig vs ROAST	Stat&Prac Sig vs ROAST	Ave Wgt State CV
ROAST	-	-	0.0015
M2K	10	10	0.0025
Region	30	30	0.0027
StateN	8	8	0.0072
StateR	8	8	0.0073
%BH10	37	2	0.0017
%BH3	0	0	0.0016
Fam8	50	1	0.0015
Fam3	50	0	0.0015

6. Limitations

Like all empirical studies, this one is subject to limitations because the available data are not completely suitable:

- The A.C.E. overestimates the A.C.E. Revision II estimate with adjustment for correlation bias, the best estimate of the true population, by about four million.
- These post-stratification designs indicate individual variables which may be helpful, but they give less insight into the results possible with logistic regression modeling with multiple variables. For example, there are preliminary estimates that in a logistic regression environment the family type variables do not drastically reduce the DSEs.
- This study does not assess small area estimates below the state level where a tract-level variable might have a greater impact on the estimates.
- In assessing utility for 2010, the general conditions will hopefully be improved over the 2000 conditions represented in this paper; specifically, duplicates in the census will be identified and removed by census operations, there will be fewer other erroneous enumerations and omissions in the census, measurement errors of correct enumeration status for census records or residence and match status for P-sample records will be reduced, and errors remaining in the census will be identified and estimated by the improved coverage measurement procedures.

7. Conclusions

This study provides several insights which may be useful in the assessment of logistic regression modeling for Census 2010. Specifically:

- The use of a region-level variable produces A.C.E. state-level estimates which are significantly different from the ROAST estimates. The decreased sample sizes and increased variance of STATE designs result in comparisons with the ROAST estimates which are not significant.
- The idea behind a tract-level variable or a family variable is that it may be useful in identifying the types of tracts or persons with coverage problems. However, except for the tract-level variable with only three categories, the results are disappointing for the national total estimate. Also, it is already understood that special census and coverage survey collection efforts are needed for tracts with high hard-to-count scores and for persons loosely attached to housing units. These variables may be more easily included in the logistic regression models without the unwanted effects on the total estimates.
- Not surprisingly this study has failed to identify any magic variables that considerably improve the explanation of coverage in the decennial census. It has shown, that once ROAST is included in the model, additional variables provide only incremental changes for national estimates. These or other variables may have a larger effect on small areas or small population domains. It appears that a region variable may be the most promising and that a tract-level variable with only a few categories may also provide some benefit. Future work will look into crossing these two variables and examining lower levels of aggregation such as counties and places.

References

U.S. Census Bureau (2004) “Accuracy and Coverage Evaluation of Census 2000: Design and Methodology”. DSSD/03-DM. Issued September 2004. U.S. Census Bureau, Washington, DC.

Griffin, R, Mule, V., and Olson, D. (2006) “Census Coverage Measurement: Initial Results of Net Error Empirical Research Using Logistic Regression”, Proceedings of the Survey Research Methods Section, American Statistical Assn, Alexandria VA.

Chart 1: National Estimates for A.C.E.

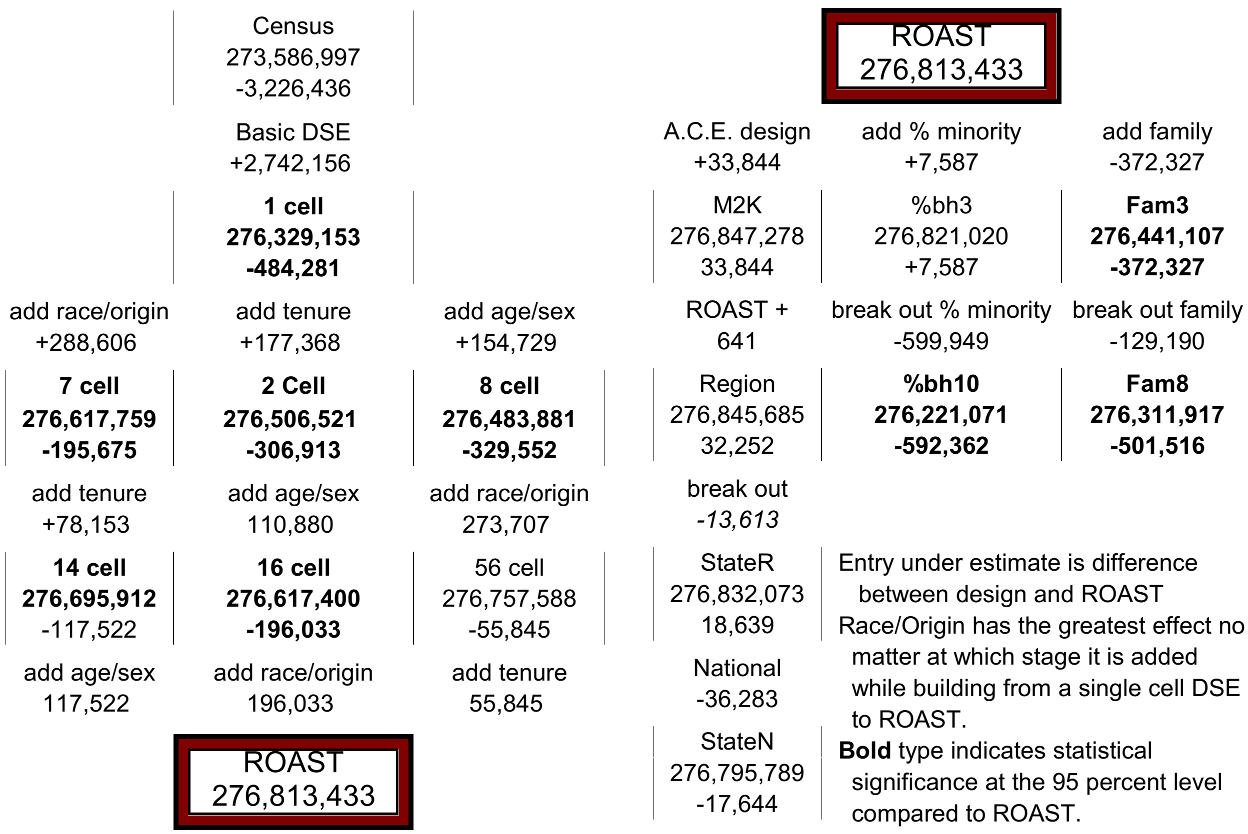

Table 2: A.C.E. Demographic Group CCFs : Significance compared to ROAST : **Statistical** *Practical* ***Both***

	Census	ROAST	M2K	%bh10	%bh3	Fam8	Fam3	Region	StateN	StateR
Total	273,586,997	1.0118	1.0119	**1.0096**	1.0118	**1.0100**	**1.0104**	1.0119	1.0117	1.0119
White	192,923,988	1.0066	1.0067	**1.0047**	1.0065	**1.0048**	**1.0054**	1.0067	1.0067	1.0067
Owner	146,957,086	1.0027	1.0029	**1.0014**	1.0027	**1.0005**	**1.0012**	1.0027	1.0023	1.0023
Renter	45,966,902	1.0191	1.0188	**1.0152**	1.0188	1.0185	1.0189	1.0197	1.0209	**1.0208**
Black	33,469,965	1.0221	1.0223	**1.0192**	1.0222	**1.0200**	1.0206	1.0222	1.0218	1.0222
Owner	16,547,598	1.0066	1.0069	**1.0045**	1.0065	**1.0041**	1.0052	1.0064	1.0052	1.0053
Renter	16,922,367	1.0373	1.0373	**1.0336**	1.0375	**1.0354**	**1.0356**	1.0377	1.0380	1.0388
Hispanic	34,538,121	1.0294	1.0294	**1.0258**	1.0297	**1.0281**	**1.0274**	1.0298	1.0290	1.0295
Owner	16,793,484	1.0121	1.0128	**1.0092**	1.0126	**1.0090**	**1.0095**	1.0118	1.0104	1.0110
Renter	17,744,637	1.0458	1.0451	**1.0416**	1.0459	1.0463	**1.0444**	1.0468	1.0466	1.0471
NHPI	590,208	1.0524	1.0524	1.0524	1.0524	1.0524	1.0524	1.0524	1.0524	1.0524
Owner	306,450	1.0280	1.0280	1.0280	1.0280	1.0280	1.0280	1.0280	1.0280	1.0280
Renter	283,758	1.0788	1.0788	1.0788	1.0788	1.0788	1.0788	1.0788	1.0788	1.0788
Asian	9,959,604	1.0097	1.0097	1.0086	1.0100	**1.0051**	1.0068	**1.0079**	1.0082	1.0081
Owner	6,032,323	1.0055	1.0055	1.0032	1.0051	**0.9998**	**1.0020**	1.0036	1.0027	**1.0025**
Renter	3,927,281	1.0160	1.0160	1.0170	1.0176	1.0132	1.0142	1.0146	1.0166	1.0168
AIR	540,158	1.0497	1.0497	1.0469	1.0505	1.0475	1.0479	1.0497	1.0480	1.0500
Owner	366,462	1.0529	1.0529	1.0513	1.0542	1.0523	1.0524	1.0529	1.0530	1.0544
Renter	173,696	1.0429	1.0429	1.0376	1.0427	1.0374	1.0386	1.0429	1.0377	1.0406
AInR	1,564,953	1.0323	1.0331	1.0290	1.0312	1.0331	1.0323	1.0323	1.0296	1.0323
Owner	921,447	1.0136	1.0149	1.0133	1.0130	1.0133	1.0136	1.0136	1.0136	1.0136
Renter	643,506	1.0590	1.0590	1.0515	1.0572	1.0616	1.0591	1.0590	1.0526	1.0590

Table 3: Sex Ratios (Significance : **Statistical** *practical* ***both***)

A.C.E.	DA Target	Census	ROAST	Effect	M2K	%BH10	%BH3	FAM8	FAM3	Region	StateN	StateR
NB 18-29 M/F	1.0437	1.0294	1.0480	129.87%	1.0481	1.0480	1.0480	**1.0502**	1.0491	1.0485	1.0477	1.0471
NB 30-49 M/F	1.0060	0.9891	0.9973	48.67%	0.9975	0.9970	0.9973	**0.9979**	**0.9977**	0.9973	0.9976	0.9977
NB 50+ M/F	0.8561	0.8445	0.8493	41.75%	0.8494	0.8491	0.8492	**0.8486**	**0.8484**	0.8493	0.8492	0.8493
B 18-29 M/F	0.8972	0.8313	0.8360	7.12%	0.8352	0.8390	0.8361	0.8395	**0.8382**	0.8369	0.8335	0.8316
B 30-49 M/F	0.8890	0.8027	0.8152	14.44%	0.8156	0.8143	0.8146	0.8168	0.8162	0.8150	0.8132	0.8126
B 50+ M/F	0.7603	0.7217	0.7223	1.47%	0.7227	0.7225	0.7224	0.7231	0.7213	0.7231	0.7230	0.7237