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

MASTER FILE

December 31, 2002

DSSD A.C.E. REVISION II MEMORANDUM SERIES # PP-14

PRED CENSUS AND SURVEY MEASUREMENT STAFF MEMORANDUM SERIES: CSM-A.C.E. REVISION II-05R

MEMORANDUM FOR: Donna Kostanich
Chair, A.C.E. Revision II Planning Group
Decennial Studies Statistics Division

From: Mary Mulry ***signed 12/31/02*** MM
Chair, A.C.E. Revision II Quality Indicators Group
Statistical Research Division

Through: David Hubble ***signed 12/31/02*** RJP for
Assistant Division Chief, Evaluations
Planning, Research, and Evaluations Division (PRED)

Prepared by: Donald Keathley
Mathematical Statistician, PRED

Subject: Specifications for Comparing Dual System Estimates for the 2000 A.C.E. Revision II Using Inmover Counts vs.Raked Outmover Counts as Estimates for the Number of Outmovers

This document contains specifications for the comparison of dual system estimates for the 2000 A.C.E. Revision II using inmover counts vs. raked outmover counts as estimates for the number of outmovers. We will use t-tests to perform these comparisons.

Contact Don Keathley at donald.h.keathley@census.gov or at 301-457-8084 if you have any questions or comments.

cc. DSSD A.C.E. Revision II Estimates Memorandum Series Distribution List
Ruth Ann Killion (PRED)
David Hubble

I. Introduction

The Accuracy and Coverage Evaluation Survey (A.C.E.) was designed to measure and possibly correct for net coverage error in Census 2000. The A.C.E., as originally designed and conducted, was fundamentally flawed. The Census Bureau's Executive Steering Committee on A.C.E. policy (ESCAP) recommended twice not to correct the census counts. There are, however, concerns about differential coverage error in Census 2000 data. While the Census 2000 data products will not be corrected, it is possible that improvements could be made to the intercensal and population estimates used for survey controls. This is the motivation for correcting errors in the A.C.E. data and developing improved estimates of the net undercount. These estimates are referred to as A.C.E. Revision II estimates. It's hoped that A.C.E. Revision II estimates will give a better picture of Census 2000 coverage and, in turn, improve operations for Census 2010 and methods for measuring coverage.

One of the possible errors in the A.C.E. is the use of inmover counts to estimate the number of outmovers. Outmovers are persons who lived at a particular address on census day but did not live at the same address during the A.C.E. interview; inmovers are persons who did not live at a particular address on census day but who did live at the address during the A.C.E. interview. This method assumes that the outmover counts from the A.C.E. underestimate the true outmover counts. This is because the A.C.E. used proxy interviews to identify outmovers and that the respondents of these interviews didn't always identify everyone who had moved away after census day.

Theoretically, the number of inmovers in the United States should equal the number of outmovers. The A.C.E. used post-stratum-level inmover counts , however, and it's possible that these counts are biased estimates of the corresponding post-stratum numbers of outmovers. In turn, this would bias the 2000 A.C.E. dual system estimates (DSE).

We're interested in learning how much of an effect the use of the inmover counts had on the DSEs. One way of doing this is to apply the distribution of outmover counts to the inmover counts using a raking procedure. The procedure would be based on various demographic and geographic variables (not necessarily the post-stratum defintions). Then we can compare the A.C.E. Revision II DSEs, by P-Sample post-stratum, FIPS state code, and region, using inmover counts vs. raked outmover counts as the estimate for the number of census movers.

We're using t-tests to compare the DSEs - this document provides the specifications for these comparisons.

II. Input Files

File(s) containing:

A. P-Sample post stratum
B. FIPS state code
C. Region
D. DSE, using inmover counts, for each P-Sample post-stratum, FIPS state code, and region
E. DSE, using raked outmover counts, for each P-Sample post-stratum, FIPS state code and region
F. each DSE's standard error

DSSD will provide these items.

III. Procedure

A. Compute differences

Compute the differences of the DSEs:

$$DSE_{difference,\ i} = DSE_{inmover,\ i} - DSE_{raked,\ i} \quad (1)$$

where $i \in$ (P-Sample post stratum, FIPS state code, region)

Example: Suppose that for a given post stratum p, the two DSEs are:

$DSE_{inmover,\ p} = 1{,}389.27$
$DSE_{raked,\ p} = 1{,}229.36$

The difference in these DSEs is:

$DSE_{difference,\ p} = 1{,}389.27 - 1{,}229.36 = 159.91$

B. Compute standard errors of the differences

Compute the standard errors of the differences of the DSEs:

$$SE(DSE_{difference,\ i}) = [\ Var(DSE_{difference,\ i})\ ]^{½}, \quad (2)$$

where $i \in$ (P-Sample post stratum, FIPS state code, region)
$Var(DSE_{difference,\ i})$ = variance of the difference in the DSEs

$$= \text{Var}(DSE_{raked,\,i}) + \text{Var}(DSE_{inmover,\,i}) - [\,2 \cdot \rho \cdot SE(DSE_{raked,\,i}) \cdot SE(DSE_{inmover,\,i})]$$

and
$\text{Var}(DSE_{raked,\,i})$ = variance of the raked DSE
$\text{Var}(DSE_{inmover,\,i})$ = variance of the inmover DSE
ρ = correlation coefficient
$SE(DSE_{raked,\,i})$ = standard error of the raked DSE
$SE(DSE_{inmover,\,i})$ = standard error of the inmover DSE

There's a lot of correlation between the raked and inmover DSEs in a given P-Sample poststratum, FIPS state code, and region. If we had time, we could develop a more sophisticated estimate of rho. Instead, we decided to use correlation coefficients $\in$ (0.90, 0.99).

Example, cont'd: Suppose that for the DSEs above, their variances are:

$\text{Var}(DSE_{inmover,\,p}) = 1{,}057.66$
$\text{Var}(DSE_{raked,\,p}) = 1{,}021.41$

Thus, the standard error of $DSE_{difference,\,p} = 159.91$ is:

$$SE(DSE_{difference,\,p}) = [\,1{,}057.66 + 1{,}021.41 - (2 \cdot 0.9 \cdot 32.52 \cdot 31.96)\,]^{1/2}$$
$$= 14.43$$

C. Compute the t-statistic for each difference

The t-statistic for each difference is:

$$t_i = DSE_{difference,\,i} \;/\; SE(DSE_{difference,\,i}) \qquad (3)$$

Example, cont'd: The t-statistic for the values above is:

$$t_p = 159.91 \;/\; 14.43$$
$$= 11.08$$

D. Determine if a difference is significant

Using a signficance level of 0.10 in a two-tailed test (α=0.10, $\alpha/2$=0.05), determine if the differences of the DSEs are signficant.

1. the t value $t_{\alpha/2, \infty}$, = 1.645 (where $\alpha/2=0.05$, DF=∞[1]) for every P-Sample post-stratum, FIPS state code, and region
2. compare $t_{\alpha/2, \infty}$ to every t_i
3. if $t_{\alpha/2, DF, i} < t_i$, then the difference is significant

Example, cont'd: From above, $t_p = 11.08$; $1.96 < t_p = 11.08$. Thus, the difference of the DSEs in post-stratum p is significant.

E. Flag each significant difference

IV. Output Files

A. Post-stratum--level file

This file is **rakevar2.sas7bdat** - it is on *dssd_se_sam:[d_olson.zuw]* on the vax. The variables in the data set are:

1. PGRP - full P-Sample post-stratum code (see Keathley (2002) for post-stratum code definitions)
2. RDSE - DSE_{raked} values
3. RVAR - standard errors (not variances) of the DSE_{raked} values
4. PDSE - $DSE_{inmover}$ values
5. PVAR - standard errors (not variances) of the $DSE_{inmover}$ values
6. DIFF - differences of the DSEs
7. SEDIFF - standard error of the differences
8. TDIFF - t-value of the differences
9. TFLAG - significance flag, where:
   - 0 = non-significant difference
   - 1 = significant difference

[1] Some of the P-Sample post-strata had sample sizes as low as 100. Letting DF=sample size, with $\alpha=0.10$, would have resulted in $t_{\alpha/2, DF}$ values being as high as 1.661. This maximum is 0.016 greater than $t_{\alpha/2, \infty} = 1.645$. Since this difference was so relatively small, combined with the fact that we were under a tight deadline, we decided to let the degrees of freedom = ∞ for all P-Sample post-strata, FIPS state codes, and regions. This, in turn, allowed the programmers to finish their work more quickly.

B. State-level file

This file is **don_state.sas7bdat** - it is on *ires$disk:[ires.rz.loss_func]* on the vax. The variables in the data set are:

1. st - FIPS state code
2. state - state name
3. d - $DSE_{inmover}$
4. se - standard error of $DSE_{inmover}$
5. d_rake - DSE_{raked}
6. se_rake - standard error of DSE_{raked}

C. Region-level file

This file is **don_region.sas7bdat** - it is on *ires$disk:[ires.rz.loss_func]* on the vax. The variables in the data set are:

1. Region
2. d - $DSE_{inmover}$
3. se - standard error of $DSE_{inmover}$
4. d_rake - DSE_{raked}
5. se_rake - standard error of DSE_{raked}

V. Reference

Keathley, Don (2002), *"Procedure for Raking Outmovers to Inmovers in the A.C.E. Revision II for the Evaluation of using Inmovers to Estimate the Number of Census Outmovers,"* DSSD A.C.E. Revision II Memorandum Series # PP-13, dated December 31, 2002