December 31, 2002

MASTER FILE

DSSD A.C.E. REVISION II MEMORANDUM SERIES #PP-34

MEMORANDUM FOR   Donna Kostanich
                        Chair, A.C.E. Revision II Planning Group

From: Mary Mulry
                        Chair, Quality Indicators Group

Prepared by: Douglas Olson
                        Sample Design Staff
                        Decennial Statistical Studies Division

Subject: A.C.E. Revision II: Overview of Variance Estimation

The variance estimation system for the Accuracy and Coverage Evaluation (A.C.E.) Revision II is a stratified jackknife generally following the methods of the A.C.E. variance estimation system described in (Kim, 2000), with suitable modifications. The system adjusts for most of the variance stemming from the block cluster sampling:

- Two-phase block cluster sampling for the full E- and P-samples
- The additional (third) level of subsampling represented by the Revision sample
- The different post-stratification systems used on the E- and P-sides

No direct attempt was made to capture the covariance structure of TES subsampling, which was minor. Also, there was no attempt to directly capture variance from any other cause, such as imputation, duplication study results, correlation bias correction or subsampling within clusters. Not estimating a variance for correlation bias correction possibly overstates the variance of some population groups (mostly adult males) which often have high erroneous enumeration and non-match rates, and understates that of aggregates that include members of both sexes, because the correlation bias correction effectively reduces the sample size by substituting female data for male.

Replication of individual samples

For purposes of understanding the replication methodology, it would be useful to think of the existence of four separate samples and consider then individually: the full E-Sample, the full P-Sample, the Revision E-Sample and the Revision P-Sample. The two full samples were

replicated using the methods described in (Kim, 2000), with the samples post-stratified by EFULLSHORT and PFULLSHORT, as described in (Haines, 2002a). The two revision samples were replicated using the method described in (Olson, 2002), post-stratified using the system described in (Haines 2002b).

The full samples contained 29,136 replicates representing each cluster in the first phase of the full sample. The revision sample contained 2,259 replicates representing each cluster selected into the sample.

Combining of samples

The 525 EFULLSHORTs and 480 PFULLSHORTs represent collapses by age and sex of the full set of 744 E- and 512 P post-strata, which cross in 7,584 ways, and are expanded to the full set for calculation of population totals. The dual system estimate (DSE) of each crossed post-stratum was calculated from the formula in (Kostanich, 2003), with an additional adjustment for correlation bias correction, expressed as a two-step process involving a coverage correction factor (CCF):

$$CCF_{hl} = (1 - r_{II}) \times C_{\kappa(hl)} \times rCE_{i(h)} \div rM_{j(k)}$$

$$DSE_{hl} = Census_{hl} \times CCF_{hl}$$

h and l are the expanded E- and P-post-strata, which were collapsed to i,j for computation of rCE and rM, which are the final correct enumeration and match rates. The r_{II} is the adjustment of the census count for data defined persons. C_k is the correction factor for correlation bias (set equal to 1 for post-strata that are not adjusted.)

In variance estimation, the rCE and rM are each assigned a replicate value for each cluster in the full sample, rCE_c and rM_c, from which a replicate value CCF_c is constructed. The principal deliverable for variance estimation was a set of 29,136 replicate $CCF_{hl,c}$ for each of the 7,584 crossed post-strata, used in the calculation of variances of roll-ups and synthetic estimates and in the construction of loss functions. A similar set of replicate CCF's was created from a set of P-sample data in which the outmovers had been adjusted by a raking procedure.

Combining of Crossed Post-strata

Because of the enormous number of crossed post-strata, all variances produced as part of the A.C.E. Revision II study represented either roll-ups of post-strata or synthetic estimates of sums of the individual post-strata. The mathematical formula for each type is the same:

$$Syn = \sum_{hl} Pop_{hl} \times CCF_{hl}$$

The difference between a roll-up and a synthetic estimate is that in a roll-up the Pop term is always either the full national Census count of household persons in the crossed post-stratum or zero, while in a synthetic estimate Pop represents the count of household persons in the post-stratum in the universe of interest, usually a geographic area. Again, to estimate the variance, perform the above summation for each replicate, substituting $CCF_{hl,c}$ into the formula to calculate a set of Syn_c, and then tabulate the squared differences among the replicate synthetic estimates from:

$$Var(Syn) = \sum_{c} K_c \times (Syn_c - Syn)^2$$

where K_c is a technical adjustment necessary to avoid overstating the variance by implicitly understating the size of the sample. K_c is equal to $(1 - 1/NISS_c)$, where $NISS_c$ is the number of clusters in the first-phase sampling stratum (of the full A.C.E. sample) to which cluster c belongs. If $NISS_c$ equals 1, K_c is assigned the value 1.

References

Kim, Jae Kwang, Alfredo Navarro and Wayne Fuller (2000); Variance Estimation for the 2000 Census Coverage Estimates. *Proceedings of the JSM 2000, p515-520*

Olson, Douglas (2002); Variance Estimation in the Census 2000 Subsamples; *Proceedings of the JSM 2002, (as yet unpublished)*

Haines, Dawn (2002a); A.C.E. Revision II Computer Specifications for Defining Full P- and E-sample Post-strata; DSSD A.C.E. REVISION II MEMORANDUM SERIES #PP-26

Haines, Dawn (2002b); A.C.E. Revision II Computer Specifications for Defining Revision P- and E-sample Measurement Error Correction Post-strata; DSSD A.C.E. REVISION II MEMORANDUM SERIES #PP-27

Kostanich, Donna (2003); A.C.E. Revision II Design and Methodology; DSSD A.C.E. REVISION II MEMORANDUM SERIES #PP-30; Chapter 6