VARIANCE ESTIMATION FOR PERSON DATA USING SUDAAN AND THE NATIONAL HEALTH INTERVIEW SURVEY (NHIS)

PUBLIC USE PERSON DATA FILES, 1994 - 1995
Combining '94 and '95 Data Only

1. Put a year ID on each data set, say YEAR *(where '94 = 1 and '95 = 2)*

2. For 1994 data, use method 1 approach, but rename variables:
   (drop the "c" on the variable names so they match with the variable names in '95 file)

   STRATUM = CSTRATUM; (pseudo psu codes; tape location 187-188)*
   PSU = CPSU; (pseudo psu codes; tape location 189)*

3. For 1995 data, use method 1 approach and the following the SAS code:
   (use 1995 SAS code to create stratum and psu variables; design variables for variance estimation might reside in different tape locations in different supplements - check for correct tape locations)

Variable Name	Tape Location*	Field Label
STRAT_V	337-340	'STRATA FOR VARIANCE ESTIMATION'
PSU_V	341	'PSU FOR VARIANCE ESTIMATION'
TYPE_PSU	351	'TYPE OF PSU'
PANEL	352	'PANEL 4'

1995 SAS® code for method 1:

```
STRATUM  = STRAT_V   ;
PSU = PANEL  ;
IF ( PSU_V = 5 ) THEN PSU = INT( ( PANEL + 1) / 2 ) ;
IF ( PSU_V = 8 ) THEN STRATUM  = 553 ;
IF ( (TYPE_PSU  = 1) AND (PSU_V IN (2,4) ) ) THEN STRATUM = (STRAT_V - 1 );
IF ( (STRAT_V  = 921)  AND (PSU_V  = 3) )  THEN STRATUM  = 901 ;
```

4. Concatenate '94 and '95 data sets together.
   (Before concatenating data files, make sure each file has the same variable names.)

5. Create a modified weight variable which will provide an estimate for an average of the two years of data (the weights then provide a mid-point estimate between the two years of data).

(*Use this SAS code*)

MWEIGHT = FABW / 2; (Final Annual Basic Weight [FABW]; tape loc. 207-212)*

6. Sort concatenated file by year, stratum, psu

7. Use SUDAAN**, where:

PROC (procedure name) DESIGN = WR;
NEST year stratum psu /psulev=3 ;
WEIGHT mweight;

(the psulev = 3 tells SUDAAN that psu is in the third position - usually it expects psu to be in the second position on the nest statement)

Note: Because the NHIS sampling design changed between 1994 and 1995, this SUDAAN design used for combining 1994 and 1995 data reads the years as independent and treats them like strata.

** These tape locations might be different in different supplements - check for correct tape locations.*

*** Two issues analyst need to be aware of in using NHIS-D, Phase 2 (DFS) data files:*

*1) If using the **1995** Adult or Child NHIS-DFS (phase 2) data files, the MISSUNIT command needs to be added to the NEST statement in SUDAAN:*

NEST = stratum psu/missunit; [using the 1995 Adult or Child NHIS-DFS only]

NEST = year stratum psu/psulev=3 missunit; [using the 1995 Adult or Child NHIS-DFS concatenated with 1994 data]

The 1995 WR design has exactly 2 PSUs per stratum and with some PSUs missing in the 1995 NHIS-DFS files, the SUDAAN MISSUNIT option performs a fix-up which produces a standard error identical to that achieved when using a full data set. Note, other output like design effects, degrees of freedom, and standardization may be computed differently. The user is responsible for checking that input leads to correct results.

*2) All NHIS-D, Phase 2 data files are **NOT** robust to subsetting.*