

## NATIONAL EMPLOYER HEALTH INSURANCE SURVEY SAMPLE SUDAAN PROCEDURE STATEMENTS

Specialized software are needed to compute sample variances from complex sample surveys such as the NEHIS. Specialized software is needed because standard statistical software packages generally do not take into account four common characteristics of sample survey data: (1) unequal probability selection of observations, (2) clustering of observations, (3) stratification, and (4) nonresponse and other adjustments. The sample weights included in each of the three NEHIS files (establishment, plan, SENE) reflect the selection probabilities of sample units and other survey design features such as stratification and clustering. Ignoring these weights will result in biased point estimates. Even when standard packages include the weight while generating point estimates, the standard errors that they generate often underestimate the estimated variance because they ignore the impact of clustering, stratification and the weights.

It is recommended that analysts use SUDAAN software to compute sampling variances for estimates based on the 1994 National Employer Health Insurance Survey files, since this software has been designed to account for design features of complex surveys. The NEHIS files reflect several samples, each of which can be used for certain types of analyses. Each sample requires different sample design statements and/or weights. Sample SUDAAN programs are provided for each of these samples.

The SUDAAN software available at the NCHS RDC is the stand-alone SUDAAN Release Version 7.5 for Windows 95. This version of SUDAAN uses TEXT files. The SUDAAN statements that follow are tailored for this version of SUDAAN. The user should take this into consideration when using these examples or parts of these examples verbatim.

The example procedures represent relatively simple, straightforward applications. The options (various parameters, test statistics, etc.) in the sample programs may not be suitable for all needs. Our intention is provide examples with the key parameters that reflect the NEHIS sample design. These parameters are found in the SUDAAN *design*, *weight*, *nest*, *totcnt* statements. In addition, the examples are limited to simple descriptive procedures for producing means or percentages. The same sample parameters for descriptive procedures are used for more complex estimation procedures such as regression or logistic regression.

Separate establishment-, plan-, and person-level examples are provided. Preprocessing or recoding may be required for some variables because of missing data or other recoding of data. Classification (SUBGROUP) variables with zero values will be treated by SUDAAN as missing and dropped from the procedure. This does not hold true for analysis variables (VAR) where zero is a valid response. Records with zero weights will automatically be excluded from the estimates produced in SUDAAN procedures. To add labels for levels of classification variables, a LEVEL.DBS file must be created. This file includes the variable name, the level of the variable and the label associated with the level of the variable.

In using SUDAAN, the full population must be processed even when analyses are for subgroups or subpopulations. This is to ensure the correct computation of the sampling variance. The SUDAAN statement SUBPOPN should be used to identify the specific analytic subpopulation of interest. If the file is reduced to a specific subpopulation prior to SUDAAN processing, the sampling variance estimates SUDAAN computes may be wrong.

### **Establishment estimates for percentage of establishments**

This example estimates the percentage of private (SAMPTYPE=1) establishments offering health insurance by firm size groups. The groups are defined in the RECODE statement. The sum of the weights, FESTB\_WT, estimates the total number of establishments. Standard errors of the percentages, weighted and unweighted population counts, and sample design effects are included in the output. Both FIRMCAT and INSURE2 include positive integer codes to conform to SUDAAN conventions for SUBGROUP variables. The SUBPOPN statement is used to identify the subpopulation of interest within the overall population.

```
PROC CROSSTAB DATA={name supplied by user} FILETYPE=sas DESIGN=strwor;
  NEST varstra2;
  TOTCNT totcnt;
  WEIGHT festb_wt;
  SUBPOPN samptype=1;
  SUBGROUP firmcat insure2;
  LEVELS 4 2;
  RECODE FIRMSIZN=(1,10,25,100);
  TABLES firmcat*insure2;
  SETENV COLWIDTH=8 DECWIDTH=2;
  PRINT NSUM = "Sample size"
        WSUM = "Number of establishments"
        SEWGT = "Standard error for establishments"
        ROWPER= "Percent"
        SEROW = "Standard error of percent"
        DEFFROW = "Design effect" /
        STYLE=NCHS NSUMFMT=F6.0 WSUMFMT=F8.0 DEFFROWFMT=F6.2;
  TITLE "Percent of private establishments offering health insurance by firm size";
```

### **Establishment file estimates of percent of private employees participating in their employers health plan as of 12/31/93.**

This example estimates a percentage of employees from the establishment file using PROC RATIO. This procedure is used because the numerator and denominator of the percentage are two separate continuous variables. The numerator is the number of employees participating in their employers health plan, with imputed values when data was missing (I\_EESCOV). The

denominator is the total number of employees working in the establishment (ESTSIZEN). The SUBPOPN statement is used to identify the private sector subpopulation of interest (SAMPTYPE=1) within the overall population. The STYLE=NCHS statement outputs the estimates and standard errors in a tabular form.

```

PROC RATIO DATA={name supplied by user} FILETYPE=sas DESIGN=strwor;
  NEST varstra2;
  TOTCNT totcnt;
  WEIGHT festb_wt;
  SUBPOPN samptype=1;
  SUBGROUP firmcat;
  LEVELS 4;
  NUMER i_eescov;
  DENOM estsizen;
  TABLES firmcat;
  SETENV COLWIDTH=8 DECWIDTH=2;
  PRINT NSUM = "Sample size"
        WYSUM = "Total employees"
        WXSUM = "Participating employees"
        RHAT = "Percent of participating employees"
        SERHAT = "Standard error"
        DEFFRHAT = "Design effect" /
  STYLE=NCHS NSUMFMT=F6.0 WYSUMFMT=F10.0 WXSUMFMT=F10.0
  RHATFMT=F6.3 SERHATFMT=F6.3
  WSUMFMT=F8.0 DEFFRHATFMT=F6.2;
  TITLE "% private employees participating in their employers' health plan by firm size";

```

**Plan file estimates of average monthly premiums paid for employees with single coverage by firm size.**

This example estimates the average premium for employees with single coverage. Since the premium desired is for comprehensive health plans (i.e., conventional indemnity, HMO, PPO, POS), the estimate is limited to these types of plans (PLANTYP4<5) in the SUBGROUP statement. The SUBGROUP statement also limits the premium estimates to private establishments (SAMPTYPE=1), and plan file records (PLANDATA=1) and cases with non-missing enrolled employees with single coverage and non-missing premiums (PLANCNT=1). PLANDATA=1 indicates plan file records (i.e., PLANDATA=0 are dummy records that were added to the plan file only for variance estimation purposes; these need to be excluded from analyses). PLANCNT=1 indicates cases where IEESINER is not missing. IEESINER may be missing because 1) all employees in the establishment had family coverage, 2) the plan was offered to employees but no employees enrolled in the plan, or 3) the establishment had plans with dual coverage (all employees were covered by both plans) and enrollment was unduplicated by setting the enrollment in one of the plans to inapplicable. PLANCNT=1 also indicates cases

where premiums or premium equivalents are not missing. Premium information was not collected for single service and special plans. The weight, NEWWT, is recomputed in the SAS program to equal NCHSPLWT\*IEESINER. NEWWT is pre-multiplied by the number of employees with single coverage so that the average premium reflects the amount paid by employees with that coverage. Since some employers reported a composite premium rate that did not differentiate between single and family coverage, the single premium is cross-tabulated by COMPIND, the indicator for composite premiums (1=composite premium, 2=premiums differ for single and family coverage).

```

NEWWT=1;
IF 0<IEESINER<9999998 THEN NEWWT=NCHSPLWT*IEESINER;
IF ISNGPREM<99998 AND 0<IEESINER<9999998 THEN PLANCNT=1;
ELSE PLANCNT=2;

PROC RATIO DATA={name supplied by user} FILETYPE=sas DESIGN=wor;
  NEST varstra2 estbno;
  TOTCNT totcnt totplan;
  WEIGHT newwt;
  SUBPOPN plandata=1 & samptype=1 & plantyp4<5 & plancnt=1 & compind=2;
  SUBGROUP firmcat;
  LEVELS 4;
  NUMER isngprem;
  DENOM plancnt;
  TABLES firmcat*compind;
  SETENV COLWIDTH=8 DECWIDTH=2;
  PRINT NSUM = "Sample size"
        RHAT = "Average monthly premium for single coverage"
        SERHAT = "Standard error of premium"
        DEFFRHAT = "Design effect of premium" /
        STYLE=NCHS NSUMFMT=F6.0
        RHATFMT=F6.3 SERHATFMT=F6.3;
  TITLE "Avg single monthly premium paid by private employees by firm size";

```

**Plan file estimates of average monthly premiums for single coverage offered to employees by firm size.**

This example, in contrast to the previous example, estimates the average premium for single coverage offered to employees. This example is almost identical to the previous example except for the WEIGHT and SUBPOPN statements. In this example, the weight statement uses the sample weight, NCHSPLWT, instead of the new weight that pre-multiplies NCHSPLWT with the number of employees enrolled in the plan. The definition of PLANCNT included in the SUBPOPN statement also differs in this example. Rather than selecting plans with premiums and enrolled employees, PLANCNT only includes cases with premiums.

```

IF ISNGPREM<99998 THEN PLANCNT=1;
ELSE PLANCNT=2;

PROC RATIO DATA={name supplied by user} FILETYPE=sas DESIGN=wor;
  NEST varstra2 estbno;
  TOTCNT totcnt totplan;
  WEIGHT nchsplwt;
  SUBPOPN plandata=1 & samptype=1 & plantyp4<5 & plancnt=1 & compind=2;
  SUBGROUP firmcat;
  LEVELS 4;
  NUMER isngprem;
  DENOM plancnt;
  TABLES firmcat*compind;
  SETENV COLWIDTH=8 DECWIDTH=2;
  PRINT NSUM = "Sample size"
    RHAT = "Average monthly premium for single coverage offered"
    SERHAT = "Standard error of premium"
    DEFFRHAT = "Design effect of premium" /
    STYLE=NCHS NSUMFMT=F6.0
    RHATFMT=F6.3 SERHATFMT=F6.3;
  TITLE "Average monthly premium for single coverage of plans offered by firm size";

```

**SENE file estimates of percent of privately insured self employed persons with no employees by source of coverage and marital status.**

This example estimates the percent of self-employed persons with no employees (SENE) under 65 years of age with health insurance coverage from 1) directly purchased plans (including insurance from the SENE's own business), 2) from the spouse's employer-sponsored health plan, 3) from the SENE's current or former employer, or 4) other sources. The sum of the weights, NRWT1, estimates the total number of SENE's under 65 years of age. Standard errors of the percentages, weighted and unweighted population counts, and sample design effects are included in the output. The SUBPOPN statement is used to identify the subpopulation of interest within the overall population (SENE's under 65 years of age with private health insurance) and does not include dummy records. Note that INSOURCE combined directly purchased health insurance (INSOURCE=1) and insurance from the SENE's business (INSOURCE=4) prior to recoding in the SUDAAN procedure. Both INSOURCE and MARSTAT were recoded to include consecutive positive integer codes to conform to SUDAAN conventions for SUBGROUP variables. Since the SUDAAN program is not SAS-callable, any recoding of non-consecutively numbered variables should be output separately for input into SUDAAN. The STYLE=NCHS statement outputs the estimates and standard errors in a tabular form.

```

If INSOURCE=4 THEN INSOURCE=1;

```

```

PROC RATIO DATA={name supplied by user} FILETYPE=sas DESIGN=wr;
NEST cstratum cpsu;
WEIGHT nrwt1;
SUBPOPN senedata=1 & ageyr<65 & privinsu=1;
RECODE insource=(1,2,3,5)
      marstat=(1,3);
SUBGROUP insource marstat;
LEVELS 4 2;
TABLES insource * marstat;
SETENV COLWIDTH=8 DECWIDTH=2;
PRINT NSUM = "Sample size"
      WSUM = "Number of SENEs"
      SEWGT = "Standard error for SENEs"
      ROWPER = "Percent"
      SEROW = "Standard error of percent"
      DEFFROW = "Design effect" /
      STYLE=NCHS NSUMFMT=F6.0 WSUMFMT=F8.0 DEFFROWFMT=F6.2;
TITLE "Percent of priv. insured SENEs <65 by source of coverage and marital status";

```