DESCRIPT Example #5

SUDAAN Statements and Results Illustrated

- CATLEVEL
- VAR
- DEFT1 option
- RLABEL
- RFORMAT

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Estimate the percentage distribution for categorical variables (arthritis, taking prescription medication, self-rated health status) for U.S. adults, using NHANES III.

Solution

The data set is comprised of adults aged 17 and older from NHANES III. (All variables in this example are from the home interview; six years of data are analyzed. The weight variable for home interview variables is WTPFQX6, and the stratification and PSU variables for analyzing six years of data are SDPSTRA6 and SDPPSU6, respectively.

- *HAC1A* is whether a doctor ever told you that you have arthritis (1=yes, 2=no);

- *HAX1B* is whether you have taken prescription medication during the past month (1=yes, 2=no); and

- *HAB1* is self-rated health status (1=excellent, 2=very good, 3=good, 4=fair, 5=poor).

Each of these variables is identified on the VAR statement below (***Exhibit 1***). The CATLEVEL statement indicates for which level(s) of each variable the percentage of adults is estimated. The first variable on the VAR statement goes with the first number on the CATLEVEL statement, and so on. For example, the percentage of adults at level (code) '1' for HAC1A is to be estimated (*i.e.,* percentage of adults with arthritis). The percentage of adults who took prescription medication last month is to be estimated (HAX1B, the second variable on the VAR statement, and '1,' the second number on the CATLEVEL statement). Finally, the percentage of adults at each of five levels of self-rated health status is to be estimated (HAB1 on the VAR statement five times, and the numbers '1' through '5' on the CATLEVEL statement). The default design effect DEFT1 is requested on the PROC statement for all estimated percentages. RLABEL and RFORMAT statements are included to label and format variables.

This example was run in SAS-Callable SUDAAN, and the programming code is presented below. Note that the basic SUDAAN code is the same for both Standalone and SAS-Callable versions.

Exhibit 1.　SAS-Callable SUDAAN Code

```
libname in "\\rtints29\sudaan\data\nhanes3";
options linesize=95 pagesize=60 nocenter;

proc format;
  value yesno 1="1=Yes";
  value health 1="1=Excel"
               2="2=Very Good"
               3="3=Good"
               4="4=Fair"
               5="5=Poor";

PROC DESCRIPT DATA=in.HANES3S3 FILETYPE=SAS DESIGN=WR DEFT1 nomarg;
  NEST SDPSTRA6 SDPPSU6;
  WEIGHT WTPFQX6;

  VAR     HAC1A HAX1B HAB1 HAB1 HAB1 HAB1 HAB1;
  CATLEVEL 1     1     1    2    3    4    5;

  SETENV ROWWIDTH=22 LBLWIDTH=23 COLWIDTH=12;
  PRINT NSUM="Sample Size" WSUM="Population Size" TOTAL SETOTAL PERCENT SEPERCENT
        LOWPCT UPPCT DEFFPCT="Deff1" / nsumfmt=f12.0 wsumfmt=f12.0 totalfmt=f12.0;

  RLABEL hac1a="Diagnosed Arthritis";
  RLABEL hax1b="Rx Meds Past Month";
  RLABEL hab1="Health Status";

  RFORMAT hac1a yesno.;
  RFORMAT hax1b yesno.;
  RFORMAT hab1 health.;

  RTITLE "PERCENTAGE OF ADULTS (17+) WITH ARTHRITIS, WHO TOOK PRESCRIPTION"
         "MEDS PAST MONTH, AND AT 5 LEVELS OF SELF-RATED HEALTH STATUS";
  RFOOTNOTE "NHANES-III, 1988-1994, JULY 1997 DATA RELEASE";
```

Exhibit 2.　First Page of SUDAAN Output (SAS *.lst file)

```
                              S U D A A N
             Software for the Statistical Analysis of Correlated Data
           Copyright    Research Triangle Institute      December 2011
                            Release 11.0.0


DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a
With Replacement (WR) Design
    Sample Weight: WTPFQX6
    Stratification Variables(s): SDPSTRA6
    Primary Sampling Unit: SDPPSU6


Number of observations read    :  20050    Weighted count :187647206
Denominator degrees of freedom :     49
```

From *Exhibit 2*, note that the total number of interviewed adults is 20,050, and each of these sample adults has a positive value for the weight variable WTPFQX6. The 20,050 sample adults make inference to a population of 187,647,206 U.S. adults aged 17 and older.

Exhibit 3. DESCRIPT Percentages

```
Variance Estimation Method: Taylor Series (WR)

PERCENTAGE OF ADULTS (17+) WITH ARTHRITIS, WHO TOOK PRESCRIPTION
MEDS PAST MONTH, AND AT 5 LEVELS OF SELF-RATED HEALTH STATUS

by: Variable, SUDAAN Reserved Variable One.
----------------------------------------------------------------------
|                       |                        | SUDAAN        |
| Variable              |                        | Reserved      |
|                       |                        | Variable One  |
|                       |                        |---------------|
|                       |                        | 1             |
----------------------------------------------------------------------
|                       |                        |               |
| Diagnosed Arthritis:  | Sample Size            |        20046  |
| 1=Yes                 | Population Size         |    187611487  |
|                       | Total                  |     32666641  |
|                       | SE Total               |   1344678.64  |
|                       | Percent                |        17.41  |
|                       | SE Percent             |         0.51  |
|                       | Lower 95% Limit Percent |        16.41  |
|                       | Upper 95% Limit Percent |        18.46  |
|                       | Deff1                  |         3.63  |
----------------------------------------------------------------------
|                       |                        |               |
| Rx Meds Past Month:   | Sample Size            |        20014  |
| 1=Yes                 | Population Size         |    187389625  |
|                       | Total                  |     82071173  |
|                       | SE Total               |   2755766.90  |
|                       | Percent                |        43.80  |
|                       | SE Percent             |         0.72  |
|                       | Lower 95% Limit Percent |        42.35  |
|                       | Upper 95% Limit Percent |        45.25  |
|                       | Deff1                  |         4.22  |
----------------------------------------------------------------------
|                       |                        |               |
| Health Status: 1=Excel | Sample Size           |        20037  |
|                       | Population Size         |    187574776  |
|                       | Total                  |     39095226  |
|                       | SE Total               |   1930091.04  |
|                       | Percent                |        20.84  |
|                       | SE Percent             |         0.70  |
|                       | Lower 95% Limit Percent |        19.48  |
|                       | Upper 95% Limit Percent |        22.28  |
|                       | Deff1                  |         5.89  |
----------------------------------------------------------------------
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE
```

Exhibit 3 is continued on the following page.

Exhibit 3.   DESCRIPT Percentages (cont'd)

```
Variance Estimation Method: Taylor Series (WR)

PERCENTAGE OF ADULTS (17+) WITH ARTHRITIS, WHO TOOK PRESCRIPTION
MEDS PAST MONTH, AND AT 5 LEVELS OF SELF-RATED HEALTH STATUS

by: Variable, SUDAAN Reserved Variable One.
----------------------------------------------------------------
|                    |                        | SUDAAN       |
| Variable           |                        | Reserved     |
|                    |                        | Variable One |
|                    |                        |--------------|
|                    |                        | 1            |
----------------------------------------------------------------
|                    |                        |              |
| Health Status: 2=Very | Sample Size         |        20037 |
| Good               | Population Size         |    187574776 |
|                    | Total                   |     57680354 |
|                    | SE Total                |   2638297.34 |
|                    | Percent                 |        30.75 |
|                    | SE Percent              |         0.68 |
|                    | Lower 95% Limit Percent |        29.40 |
|                    | Upper 95% Limit Percent |        32.14 |
|                    | Deff1                   |         4.37 |
----------------------------------------------------------------
|                    |                        |              |
| Health Status: 3=Good | Sample Size         |        20037 |
|                    | Population Size         |    187574776 |
|                    | Total                   |     61294363 |
|                    | SE Total                |   1846359.93 |
|                    | Percent                 |        32.68 |
|                    | SE Percent              |         0.70 |
|                    | Lower 95% Limit Percent |        31.29 |
|                    | Upper 95% Limit Percent |        34.10 |
|                    | Deff1                   |         4.45 |
----------------------------------------------------------------
|                    |                        |              |
| Health Status: 4=Fair | Sample Size         |        20037 |
|                    | Population Size         |    187574776 |
|                    | Total                   |     23587804 |
|                    | SE Total                |   1223801.54 |
|                    | Percent                 |        12.58 |
|                    | SE Percent              |         0.56 |
|                    | Lower 95% Limit Percent |        11.49 |
|                    | Upper 95% Limit Percent |        13.75 |
|                    | Deff1                   |         5.78 |
----------------------------------------------------------------
|                    |                        |              |
| Health Status: 5=Poor | Sample Size         |        20037 |
|                    | Population Size         |    187574776 |
|                    | Total                   |      5917030 |
|                    | SE Total                |    378850.43 |
|                    | Percent                 |         3.15 |
|                    | SE Percent              |         0.16 |
|                    | Lower 95% Limit Percent |         2.84 |
|                    | Upper 95% Limit Percent |         3.50 |
|                    | Deff1                   |         1.77 |
----------------------------------------------------------------
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE
```

Arthritis: From *Exhibit 3,* 4 persons did not answer the arthritis question (20,050 – 20,046). The 20,046 sample adults who answered this question make inference to an adult population of 187,611,487 who would answer the question if asked. Effectively, this number represents the U.S. adult, non-institutionalized civilian population. The estimated total number of adults with arthritis (diagnosed by a

doctor) is 32,666,641, with an estimated standard error of 1,344,679. The estimated prevalence of arthritis among U.S. adults is 17.4%, with an estimated standard error of 0.51%. The 17.4% is calculated by the ratio of two estimated totals: the estimated number of adults with arthritis (32,666,641) and the estimated number of adults (187,611,487). All estimates refer to the time period 1988-1994, or to the midpoint of this interval (*i.e.*, 1991).

Prescription Medications: From *Exhibit 3*, an estimated 43.8% of U.S. adults took at least one prescription medication during the past month, with an estimated standard error of 0.72%. A 95% confidence interval on this population parameter is 43.80% +/- 2.009 (0.72%) or (42.4%, 45.3%). The figure 2.009 is from the Student *t*-distribution with 49 df, the denominator degrees of freedom for this survey.

Health Status: From *Exhibit 3*, 13 persons did not answer the self-rated health question (20,050-20,037). The estimated percentage distribution of self-rated health among U.S. adults is: 20.84% (excellent), 30.75% (very good), 32.68% (good), 12.58% (fair) and 3.15% (poor). A 95% confidence interval on the number of U.S. adults with poor health status is 5,917,030 +/- 2.009 (378,850); or 5,917,030 +/- 761,110; or, rounded to significant digits, (5,156,000 to 6,678,000).

Design Effects: All design effects are for the estimated percentage, as indicated in the PRINT statement. The default design effect DEFT1, is used here (*i.e.*, the NHANES III sampling plan is compared to a simple random sample of 20,050 from the population of adults aged 17 and over). Most design effects are large, ranging from two to six. The design effects are increased by the clustered design of NHANES III, as well as by the variability in the sampling weights due to oversampling certain subpopulations, unequal probability sampling in addition to oversampling, and weighting adjustments for unit nonresponse.