DESCRIPT Example #2

SUDAAN Statements and Results Illustrated

- PERCENTILE
- VAR
- WEIGHT
- SUBPOPN
- RFORMAT

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Estimate percentiles of body mass index for adults, by age group, using NHANES III.

Solution

The target population is the civilian, non-institutionalized population of the United States aged two months and older. In this example, we are interested in adults only, aged 17 years or older. Note that the NHANES III data set can be subsetted to adults only without losing aspects of the sampling design (*i.e.*, without losing any "pseudo-PSUs").

The analysis (dependent) variable on the VAR statement in ***Exhibit 1*** is body mass index (BMPBMI_R). PERCENTILE statement requests specific percentiles to be estimated for all variables on the VAR statement. CLASS statement identifies age group as a categorical variable, TABLES statement requests estimated percentiles by the values of this variable. The SUBPOPN statement restricts the analysis to those aged 20 years and older. The SETENV statement is used to get all columns of the printout table on one screen (or piece of paper).

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 sex 1="1=male"
            2="2=female";
  value age 1="1=17-34"
            2="2=35-49"
            3="3=50-64"
            4="4=65-90+";
  value race 1="1=nH_white"
             2="2=nH_black"
             3="3=Mex_Amer"
             4="4=other";

PROC DESCRIPT DATA=in.HANES3S3 FILETYPE=SAS DESIGN=WR;
  NEST SDPSTRA6 SDPPSU6;
  WEIGHT WTPFHX6;

  SUBPOPN HSAGEIR > 19 / NAME="ADULTS AGED 20+";

  CLASS AGEGRP4 / DIR=DESCENDING;
  TABLES AGEGRP4;

  VAR BMPBMI R;
  PERCENTILE 90 / MEDIAN QUARTILES;

  SETENV LABWIDTH=11 COLWIDTH=11 ROWWIDTH=11 COLSPCE=1;
  PRINT NSUM="SAMSIZE" QTILE SEQTILE LOWQTILE UPQTILE;
  RFORMAT agegrp4 age.;
  RTITLE "SELECTED PERCENTILES OF BMI BY AGE, U.S. CIVILIAN"
         "NON-INSTITUTIONALIZED POPN. AGED 20 YEARS OR OLDER";
  RFOOTNOTE "NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, BMI UPDATED";
```

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: WTPFHX6
    Stratification Variables(s): SDPSTRA6
    Primary Sampling Unit: SDPPSU6

Number of observations read    :  18162    Weighted count :187513911
Number of observations skipped :   1888
(WEIGHT variable nonpositive)
Observations in subpopulation  :  17030    Weighted count :177180671
Denominator degrees of freedom :     49
```

See ***Example 2*** for a discussion of the above printout.

Exhibit 3. Frequencies for CLASS Variable AGEGRP4

```
Frequencies and Values for CLASS Variables
-------------------------------------
AGEGRP4          Frequency       Value
-------------------------------------
Ordered
  Position:
  1                   4495    4=65-90+
Ordered
  Position:
  2                   3098     3=50-64
Ordered
  Position:
  3                   4180     2=35-49
Ordered
  Position:
  4                   5257     1=17-34
-------------------------------------
```

In ***Exhibit 3***, SUDAAN provides a frequency of the variable AGEGRP4 across the entire 17,030 observations on in the subpopulation defined on the SUBPOPN statement, which includes the 1,888 records with a non-positive WEIGHT variable. These frequencies are provided since the variable AGEGRP4 was specified on the CLASS statement. The values for AGEGRP4 appear in all output in descending order as a result of the DIR=DESCENDING option on the CLASS statement.

Exhibit 4. DESCRIPT Results: Percentile Estimates

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: ADULTS AGED 20+

SELECTED PERCENTILES OF BMI BY AGE, U.S. CIVILIAN
NON-INSTITUTIONALIZED POPN. AGED 20 YEARS OR OLDER

by: Variable, AGEGRP4, Percentiles.
for: Variable = Body Mass Index.
```

AGEGRP4 Percentiles	SAMSIZE	Quantile	SE Quantile	Lower 95% Limit	Upper 95% Limit
Total					
25.00	16969	22.57	0.08	22.40	22.73
50.00	16969	25.52	0.10	25.35	25.75
75.00	16969	29.38	0.14	29.08	29.63
90.00	16969	33.68	0.24	33.18	34.16
4=65-90+					
25.00	4461	22.89	0.17	22.62	23.31
50.00	4461	26.01	0.13	25.80	26.33
75.00	4461	29.51	0.15	29.18	29.77
90.00	4461	32.94	0.31	32.27	33.51
3=50-64					
25.00	3087	24.20	0.12	23.95	24.45
50.00	3087	27.08	0.15	26.80	27.41
75.00	3087	31.02	0.17	30.70	31.37
90.00	3087	35.29	0.33	34.57	35.88
2=35-49					
25.00	4174	22.91	0.16	22.67	23.30
50.00	4174	25.86	0.18	25.40	26.14
75.00	4174	29.71	0.23	29.27	30.18
90.00	4174	34.34	0.55	33.44	35.66
1=17-34					
25.00	5247	21.57	0.09	21.40	21.76
50.00	5247	24.20	0.10	23.97	24.38
75.00	5247	27.70	0.17	27.29	27.98
90.00	5247	32.02	0.31	31.57	32.80

```
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, BMI UPDATED
```

Exhibit 4 indicates that the estimated <u>median</u> (50th percentile) BMI for U.S. adults 20 years and older is 25.52, with an estimated standard error of 0.10. The 95% confidence interval on this population parameter is (25.35, 25.75). An estimated 10% of U.S. adults aged 20 years and older have a BMI above 33.68 (90th percentile). The estimated quartiles for BMI seem to increase with age until the oldest age category, then decrease.