DESCRIPT Example #1

SUDAAN Statements and Results Illustrated

- SUBPOPN
- NEST
- RFORMAT
- LEVELS
- WEIGHT

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Estimate mean body mass index for adults by sex, gender and age, 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 value of body mass index (BMPBMI) on the July 1997 ASCII data release was incorrect for a few adults and was updated on the NCHS web site. The body mass index variable analyzed here (BMPBMI_R) uses the web-corrected values (as of October 2000) for body mass index.

Height and weight, the components of body mass index, were measured at the MEC exam and at the home exam for adults who could not come to the MEC trailer. However, the home height and weight were measured only for adults aged 20 years or older. Thus, we use the weight variable WTPFHX6, which refers to exam variables measured in both the MEC and the home. We use the SUBPOPN statement to restrict the analysis to adults aged 20 and over, since adults aged 17-19 did not have height and weight measured in the home exam. Since we analyze all six years of data, NEST statement are SDPSTRA6 and SDPPSU6, respectively.

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. The PRINT statement in ***Exhibit 1*** changes the keyword NSUM to SAMSIZE and the keyword DEFFMEAN to DEFFM1 on the printed results.

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 DEFT1 CONF_LIM=95;
  NEST   SDPSTRA6 SDPPSU6;
  WEIGHT WTPFHX6;

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

  SUBGROUP HSSEX AGEGRP4 DMARETHN;
  LEVELS   2     4       4;
  TABLES HSSEX AGEGRP4 DMARETHN;

  VAR BMPBMI_R;

  PRINT  NSUM="SAMSIZE" MEAN SEMEAN DEFFMEAN="DEFFM1" LOWMEAN UPMEAN /STYLE=NCHS;
  rformat hssex sex.;
  rformat agegrp4 age.;
  rformat dmarethn race.;
  RTITLE "MEAN BMI, BY SEX, BY RACE/ETHNICITY AND BY AGE"
         "U.S.CIVILIAN, NONINSTITUTIONALIZED 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
```

The 20,050 adults in the SAS data set (18,162 + 1,888) were all interviewed at home (***Exhibit 2***). Most (18,162) were examined at the MEC or at home; 1,888 had no physical exam and thus have a value of 0 for the weight variable WTPFHX6. The sum of the values of WTPFHX6 over the 18,162 examined adults is 187,513,911, an estimate of the number of civilian, noninstitutionalized adults (17 years and older) in the U.S. at the time of NHANES III. Among the 18,162, there are 17,030 who are in the subpopulation aged 20 years and older; they represent an estimated 177,180,671 adults in the population.

denominator degrees of freedom (ddf) for any sample survey is the number of PSUs minus the number of strata. For NHANES III (using six years of data), the ddf is 98 "pseudo-PSUs" minus 49 "pseudo-strata" (i.e., 49 ddf). DESIGN=WR on the PROC statement describes with replacement sampling at the first stage [the PSUs] of NHANES III, even though the actual first-stage sampling in NHANES III was without replacement. WR is a commonly used approximation when a small percentage of PSUs are selected from every stratum. DESIGN=WR also indicates that Taylor Series linearization is to be used for variance estimation.

Exhibit 3. DESCRIPT Results: By Sex

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

MEAN BMI, BY SEX, BY RACE/ETHNICITY AND BY AGE
U.S.CIVILIAN, NONINSTITUTIONALIZED POPN. AGED 20 YEARS OR OLDER

by: Variable, Sex.
--------------------------------------------------------------------------------
Variable
   Sex                                                    Lower 95%    Upper 95%
                   SAMSIZE       Mean    SE Mean    DEFFM1   Limit Mean   Limit Mean
--------------------------------------------------------------------------------
Body Mass Index
   Total             16969      26.51       0.11      6.34        26.29        26.73
   1=male             7933      26.58       0.11      4.10        26.37        26.80
   2=female           9036      26.44       0.16      5.48        26.13        26.76
--------------------------------------------------------------------------------
NHANES-III, 1988-1994, JULY,1997 DATA RELEASE, BMI UPDATED
```

The results presented in ***Exhibit 3*** indicate that the estimated mean BMI for adults aged 20 years and older is 26.51, with an estimated standard error of 0.11. A 95% confidence interval on mean body mass index for the subpopulation (20 years and older) is 26.51 +/-2.009 × (0.11), or (26.3, 26.7). Males have a slightly higher estimated mean BMI (26.58) than do females (26.44), although the difference is small compared to the estimated standard errors. The 95% confidence limits for males and females are overlapping.

Note that 17,030 sample adults are in the subpopulation, but only 16,969 of them are in ***Exhibit 3***. This occurs because 61 (17,030 – 16,969) subjects have a missing value for BMPBMI_R (no subject has a missing value for gender, race/ethnicity, or age). The interpretation of the point estimates and the confidence interval above assumes that the values for BMPBMI_R are missing at random, given the covariates used to adjust for nonresponse in NHANES III. These covariates can generally be assumed to be gender, race/ethnicity, age, and geographic region of residence. See the NHANES III documentation for more detail on their nonresponse adjustment methods. If a large percentage of subjects do not answer a question, the inference might be better restricted to the population who would have answered the question or submitted to the measurement.

In ***Exhibit 3***, design effect of 6.34 is for the estimated mean for the total subpopulation (20 years and older). The default design effect is used here (DEFT1), and it compares the variance of the estimated mean using the NHANES III design to the variance for the estimated mean that would have been obtained with a simple random sample of 17,030 adults (less the number of adults with item nonresponse). The design effect 6.34 is increased beyond 1.0 due to clustering of adults within a PSU and by variability in the sampling weights (due to oversampling certain subpopulations, unequal probability sampling, and weighting adjustments for unit nonresponse).

Note that the design effect for the estimated mean is lower for males and females than for the entire subpopulation. This occurs because the design effect is positively associated with the average number of

subjects per PSU, and this average is smaller for males only or for females only than for the total subpopulation.

Exhibit 4. DESCRIPT Results: By Age

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

MEAN BMI, BY SEX, BY RACE/ETHNICITY AND BY AGE
U.S.CIVILIAN, NONINSTITUTIONALIZED POPN. AGED 20 YEARS OR OLDER

by: Variable, AGEGRP4.
--------------------------------------------------------------------------------
Variable
   AGEGRP4                                                  Lower 95%    Upper 95%
                  SAMSIZE     Mean    SE Mean    DEFFM1    Limit Mean   Limit Mean
--------------------------------------------------------------------------------
Body Mass Index
   Total            16969    26.51       0.11      6.34         26.29        26.73
   1=17-34           5247    25.29       0.12      3.15         25.04        25.54
   2=35-49           4174    26.97       0.19      5.49         26.58        27.36
   3=50-64           3087    27.99       0.14      1.93         27.71        28.27
   4=65-90+          4461    26.61       0.11      1.39         26.38        26.84
--------------------------------------------------------------------------------
NHANES-III, 1988-1994, JULY,1997 DATA RELEASE, BMI UPDATED
```

In ***Exhibit 4*** the estimated mean body mass index seems to increase with age until the oldest age group, where there is a decrease. The somewhat smaller design effects for older ages compared to younger ages likely occur because older persons were oversampled for NHANES III. When the comparison simple random sampling (SRS) plan is considered for the calculation of the design effect for older ages, the SRS sample size for older people will be smaller than the NHANES III sample size for older persons.

Exhibit 5. DESCRIPT Results: By Race/Ethnicity

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

MEAN BMI, BY SEX, BY RACE/ETHNICITY AND BY AGE
U.S.CIVILIAN, NONINSTITUTIONALIZED POPN. AGED 20 YEARS OR OLDER

by: Variable, Race-ethnicity.
--------------------------------------------------------------------------------
Variable
   Race-ethnicity                                           Lower 95%    Upper 95%
                  SAMSIZE     Mean    SE Mean   DEFFM1    Limit Mean   Limit Mean
--------------------------------------------------------------------------------
Body Mass Index
   Total            16969    26.51       0.11     6.34         26.29        26.73
   1=nH_white        7110    26.34       0.13     7.58         26.07        26.61
   2=nH_black        4692    27.75       0.15     0.92         27.45        28.04
   3=Mex_Amer        4499    27.41       0.15     0.68         27.11        27.71
   4=other            668    25.87       0.30     4.21         25.28        26.47
--------------------------------------------------------------------------------
NHANES-III, 1988-1994, JULY,1997 DATA RELEASE, BMI UPDATED
```

Exhibit 5 indicates that the estimated mean BMI for non-Hispanic blacks and for Mexican-Americans seems to be higher than the estimated mean BMI for non-Hispanic whites and the "other" race/ethnicity group. The 95% confidence limits for non-Hispanic blacks and Mexican Americans are also non-

overlapping with those for non-Hispanic whites and "other" race, indicating statistically significant differences between these groups.

The very low design effects for Mexican-Americans (less than 1.0 in ***Exhibit 5*** results from dramatic oversampling of this group; that is, the complex design has a much larger sample size for this group than would have been obtained under simple random sampling of adults. Specifically, Mexican-Americans comprise 26% of the NHANES III adult sample, but only 5% of the U.S. adult population. Similarly, the undersampling of non-Hispanic whites results in a large design effect (7.58), one that is larger than the design effect for the entire subpopulation (6.34). The complex design has a much smaller sample size of this race/ethnic group than would have been obtained under simple random sampling of adults. Non-Hispanic whites comprise just 42% of the NHANES III adult sample, but represent 76% of the U.S. adult population.