VARGEN Example #2

SUDAAN Statements and Results Illustrated

- XMEAN
- XRATIO
- XSUM
- PARAMETER
- SUBPOPX

Input Data Set(s): DA32722P1.SAS7bdat

Example

Using NSDUH 2010 data, this example will estimate the mean age that adults over 21, that have had a drink of alcohol, had their first drink. This estimate is created three different, yet equivalent ways.

Solution

The data set is from the National Survey on Drug Use and Health (NSDUH) in 2010. The target population of NSDUH is the non-institutionalized civilian population of United States citizens aged twelve and older. In the public use data file there are 57,873 individual observations, which is the dataset used in this example.

The variable of interest in this example is ALCTRY. ALCTRY is the age that an individual first had a drink of alcohol, where a drink is classified as an entire drink excluding a sip or a taste. In the dataset used in this example, 70.19% of respondents gave a valid age for ALCTRY. 28.72% of respondents had never had a drink of alcohol (ALCTRY = 991) and the remaining 1.1% of individuals in the sample gave bad data (ALCTRY = 985), did not know (ALCTRY = 994), refused to answer (ALCTRY = 997), or left the question blank (ALCTRY = 998). All of these values of ALCTRY that are greater than 100 will be excluded from the analysis. This example uses XMEAN, XRATIO, and a combination of XSUM statements and a PARAMETER statement to estimate the mean of ALCTRY.

An XSUM statement is used to first compute an estimate of the sum of ALCTRY, naming the variable ALCSUM. Another XSUM statement is then used again to compute an estimate of the total number of individuals in the subpopulation, naming the variable SBPOPLN. Finally, a PARAMETER statement is used to divide ALCSUM by SBPOPLN to give the mean age individuals in the subpopulation first had a drink of alcohol, this variable is called MEAN1.

The XRATIO statement computes the average the ratio of ALCTRY to 1. So essentially it is just computing the mean of ALCTRY. The variable created is called MEAN2.

The XMEAN statement directly computes the mean of ALCTRY, naming the variable MEAN3.

As will be presented in the following exhibits, MEAN1 = MEAN2 = MEAN3, and thus all three methods of x-statements are valid ways to compute the same mean. This demonstrates VARGEN's capabilities to compute statistics through numerous computation methods, and to include all the methods in the same procedure call.

The SETENV statement is used to format the output properly so that it can be correctly viewed with all numbers having enough column width to be fully displayed.

For this example we are interested in all individuals of a legal drinking age in the United States, thus persons aged 21 or older. AGE2 is the variable in the NSDUH dataset that indicates age of respondent. If AGE2 is equal to 10 then the individual was 21 years old at the time of response. By using a SUBPOPX statement the subpopulation is restricted to all respondents 21 years of age or older (AGE2 greater than or equal to 10). In addition, restrictions needed to be made to the ALCTRY variable. The subpopulation was limited to only respondents who gave a valid age for when they first had a drink of alcohol. Thus using the same SUBPOPX statement as mentioned earlier, the subpopulation was restricted to only include those who had had a drink and valid age value for ALCTRY (ALCTRY less than 100).

The variable containing the sample weights is ANALWT_C. VESTR is the analysis stratum and VEREP is the analysis replicate. Both VESTR and VEREP are used in the NEST statement. The dataset needs to be sorted by the variables in the NEST statement prior to calling VARGEN.

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

Exhibit 1.   SAS-Callable SUDAAN Code (DESIGN=WR)

```
libname lib '\\rtints29\sudaan\Data\NSDUH 2010\';
filename in '\\rtints29\sudaan\Data\NSDUH 2010\32722-0001-Data.stc';

proc cimport infile=in library=lib;
run;

data NSDUH;
set lib.da32722p1;
options fmtsearch = (lib.formats);
run;

proc sort data=NSDUH;
by vestr verep;
run;

proc vargen data=NSDUH design = wr;
weight analwt_c;
nest vestr verep;
subpopx 10 <= age2 and alctry <= 100;
xsum alcsum : alctry;
xsum sbpopln : _one_ / name = "# of Individuals in Subpopulation";
parameter mean1 : alctry / sbpopln / name = "MEAN1: Avg of ALCTRY Using XSUM";
xratio mean2 : alctry , 1 / name = "MEAN2: Avg of ALCTRY Using XRATIO";
xmean mean3 : alctry / name = "MEAN3: Avg of ALCTRY Using XMEAN";
setenv colwidth = 15;
rtitle "Estimating the Average Age of First Alcohol Try Using XSUM, XMEAN, and XRATIO";
rfootnote "Data from NSDUH 2010";
run;
```

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        June 2012
                             Release 11.0.0-testing-221


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


Number of observations read   :  57873    Weighted count :253619107
Observations in subpopulation :  28174    Weighted count :188237618
Denominator degrees of freedom :     60
```

The results displayed in *Exhibit 2* indicate that there are 28,174 sample members in the subpopulation of individuals aged 21 and older with valid values for ALCTRY as specified by the SUBPOPX statement in *Exhibit 1*. They represent 188,237,618 individuals in the target population.

Exhibit 3.　　XSUM of ALCTRY Output

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: 10 <= AGE2 AND ALCTRY <= 100
Estimating the Average Age of First Alcohol Try Using XSUM, XMEAN, and XRATIO
by: Variable, SUDAAN Reserved Variable One.


-------------------------------------------------------------------------------------------
Variable
   SUDAAN Reserved
     Variable One         Sample Size    Weighted Size        Estimate        SE Estimate
-------------------------------------------------------------------------------------------
ALCSUM
   Total                       28174     188237618.47     3271646762.99       42264601.20
   1                           28174     188237618.47     3271646762.99       42264601.20
-------------------------------------------------------------------------------------------
Data from NSDUH 2010
```

The results in *Exhibit 3* display the output from the x-statistic generated by the XSUM statement on the ALCTRY variable. As can been seen, the estimate for the sum of ALCTRY for the subpopulation is 3,271,646,762.99 with a standard error of 42,264,601.20. Note that no NAME option was used in the XSUM statement so that the created variable name, ALCSUM, appears in the output, instead of a label in its place.

Exhibit 4.　　XSUM to count Subpopulation

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: 10 <= AGE2 AND ALCTRY <= 100
Estimating the Average Age of First Alcohol Try Using XSUM, XMEAN, and XRATIO
by: Variable, SUDAAN Reserved Variable One.


-------------------------------------------------------------------------------------------
Variable
   SUDAAN Reserved
     Variable One         Sample Size    Weighted Size        Estimate        SE Estimate
-------------------------------------------------------------------------------------------
# of Individuals in
  Subpopulation
   Total                       28174     188237618.47      188237618.47        2326206.74
   1                           28174     188237618.47      188237618.47        2326206.74
-------------------------------------------------------------------------------------------
Data from NSDUH 2010
```

The results in *Exhibit 4* display the output from the XSUM statement used to estimate the subpopulation total. The estimate is that there are 188,237,618.47 individuals in the subpopulation with a standard error of 2,326,206.74. As can be seen, the estimate is the same as the weighted size. This is expected and verifies the estimate's correctness.

Exhibit 5.   PARAMETER Statement to Estimate Mean of ALCTRY

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: 10 <= AGE2 AND ALCTRY <= 100
Estimating the Average Age of First Alcohol Try Using XSUM, XMEAN, and XRATIO
by: Variable, SUDAAN Reserved Variable One.


------------------------------------------------------------------------------------------
Variable
  SUDAAN Reserved
    Variable One          Sample Size     Weighted Size          Estimate        SE Estimate
------------------------------------------------------------------------------------------
MEAN1: Avg of ALCTRY
  Using XSUM
   Total                      28174      188237618.47              17.38               0.05
   1                          28174      188237618.47              17.38               0.05
------------------------------------------------------------------------------------------
Data from NSDUH 2010
```

The estimate in *Exhibit 5* is generated by a PARAMETER statement using the XSUM estimates displayed in *Exhibit 3* and *Exhibit 4.* The estimate, called MEAN1, is result of dividing the estimate for the sum of ALCTRY for the subpopulation (ALCSUM) by the estimate for the total amount of individuals in the subpopulation (sbpopln). The MEAN1 for the subpopulation is 17.38 years old with a standard error of 0.05.

Exhibit 6.   XRATIO to Estimate Mean of ALCTRY

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: 10 <= AGE2 AND ALCTRY <= 100
Estimating the Average Age of First Alcohol Try Using XSUM, XMEAN, and XRATIO
by: Variable, SUDAAN Reserved Variable One.

-------------------------------------------------------------------------------------------
Variable
   SUDAAN Reserved
     Variable One        Sample Size      Weighted Size        Estimate      SE Estimate
-------------------------------------------------------------------------------------------
MEAN2: Avg of ALCTRY
  Using XRATIO
   Total                     28174      188237618.47             17.38            0.05
   1                         28174      188237618.47             17.38            0.05
-------------------------------------------------------------------------------------------
Data from NSDUH 2010
```

The results in *Exhibit 6* are generated using a single XRATIO statement to estimate the mean of ALCTRY for the subpopulation. This was done by finding the ratio of ALCTRY to 1. The estimate, called MEAN2, is 17.38 years of age and its standard error is 0.05.

Exhibit 7.   XMEAN to Estimate Mean of ALCTRY

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: 10 <= AGE2 AND ALCTRY <= 100
Estimating the Average Age of First Alcohol Try Using XSUM, XMEAN, and XRATIO
by: Variable, SUDAAN Reserved Variable One.

-------------------------------------------------------------------------------------------
Variable
   SUDAAN Reserved
     Variable One        Sample Size      Weighted Size        Estimate      SE Estimate
-------------------------------------------------------------------------------------------
MEAN3: Avg of ALCTRY
  Using XMEAN
   Total                     28174      188237618.47             17.38            0.05
   1                         28174      188237618.47             17.38            0.05
-------------------------------------------------------------------------------------------
Data from NSDUH 2010
```

The results in *Exhibit 7* display the output from the XMEAN statement used to estimate the mean of ALCTRY for the subpopulation. The estimate of the subpopulation mean, called MEAN3, is 17.38 years of age with a standard error of 0.05.

As can be seen by the results shown in *Exhibit 5, Exhibit 6,* and *Exhibit 7,* MEAN1, MEAN2, and MEAN3 are all estimates of the subpopulation mean of ALCTRY, and are all equal to each other. That is, MEAN1 = MEAN2 = MEAN3 = 17.38 and have identical standard errors which demonstrates that all three methods of estimating the mean are valid methods as they generate the same results. While the XMEAN statement may be the most straight forward method to do so, XSUM statements coupled with a PARAMETER statement and an XRATIO statement can be used to estimate the population mean. All three methods can be utilized in the same VARGEN procedure call.