VARGEN Example #4

SUDAAN Statements and Results Illustrated

- XRATIO
- XMEAN
- PARAMETER
- SUBPOPX
- NOTSORTED

Input Data Set(s): DA32722P1.SAS7bdat

Example

Using NSDUH 2010 data, this example looks at the subpopulation that has had a drink of alcohol in the past 12 months (at the time of the survey response) and estimates the average amount of days per week an individual has at least one drink of alcohol. This estimate is created two 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 from the dataset used to generate the estimate of days per week with at least one drink is ALCYRTOT. ALCYRTOT is the total number of days an individual has had a drink of alcohol in the past 12 months. In the dataset used in this example, 59.54% of respondents responded with a valid age for the variable (range: 1-365). 28.72% of respondents had never had a drink of alcohol (ALCYRTOT = 991) and 10.31% of the respondents had not used alcohol in the past twelve months (ALCYRTOT = 993). The remaining 1.42% of individuals in the sample gave bad data (ALCYRTOT = 985), did not know (ALCYRTOT = 994), refused to answer (ALCYRTOT = 997), or left the question blank (ALCYRTOT = 998). All observations with values of ALCYRTOT greater than 365 will be excluded from the analysis.

The example uses a SUBPOPX statement to limit the subpopulation to only include the observations with a value of ALCYRTOT of 365 or less. This is done to make the results more readily interpretable.

A CLASS statement is used to display the analysis by age using the variable CATAG6. CATAG6 has 6 levels. When CATAG6 equals 1 then the respondent is 12-17 years old, CATAG6 = 2 then 18-25 years old, CATAG6 = 3 then 26-34 years old, CATAG6 = 4 then 35-49 years old, CATAG6 = 5 then 50-64 years old, and when CATAG6 = 6 then the respondent is 65 years old or older.

An XRATIO statement is used to compute the ratio of ALCYRTOT to 52 (the number of weeks in a year). The result is called EST1, which estimates the average amount of days a week an individual has at least one drink.

An XMEAN statement is also used to compute the average of ALCYRTOT, called DAYYR. DAYYR is then divided by 52 through a PARAMETER statement to calculate EST2, which estimates the average amount of days per a week an individual has at least one drink.

A second PARAMETER statement will be used calculate the difference between EST1 (created in an XRATIO statement) and EST2 (created in a PARAMETER statement). As will be presented in the following exhibits, EST1 = EST2, which demonstrates that both methods are valid ways to estimate the average number of days in a week people have a drink(s) of alcohol. This demonstrates VARGEN's capabilities to compute the same statistics through numerous computation methods, and to compute all the methods simultaneously 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.

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.

A NOTSORTED statement is utilized which, by being included, does not require the input dataset to be sorted by the variables in the NEST statement.

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 vargen data=NSDUH design = wr NOTSORTED;
weight analwt_c;
nest vestr verep;
subpopx alcyrtot <= 365;
class catag6;
xratio est1 : alcyrtot , 52 / name="Using RATIO: Days with Drink per Week";
xmean dayyr : alcyrtot / name="Mean Drinks per Year" ;
parameter est2 : dayyr / 52 /name="Using MEAN: Days with Drink per Week";
parameter proof : est1 - est2 / name="Difference between EST1 and EST2";
setenv colwidth = 15;
rtitle1 "Estimating the Average Number of Days with at Least One Drink per Week";
rtitle2 "Done Using XRATIO and XMEAN with a PARAMETER Statement";
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      July 2012
                          Release 11.0.0-testing-233


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 :  34458   Weighted count :166239053
Denominator degrees of freedom :    60
```

The results displayed in *Exhibit 2* indicate that there are 34,458 sample members in the subpopulation of individuals who have had a drink in the last twelve months as specified by the SUBPOPX statement in *Exhibit 1*. They represent 166,239,053 individuals in the target population.

Exhibit 3.   CLASS Variable Frequencies

```
Frequencies and Values for CLASS Variables
by: AGE CATEGORY RECODE (6 LEVELS).

------------------------------------------------
AGE CATEGORY

  RECODE (6
  LEVELS)       Frequency             Value
------------------------------------------------
Ordered
  Position:
  1                5143    (1) 12-17 Years Old
Ordered
  2               15030    (2) 18-25 Years Old
Ordered
  Position:
  3                4634    (3) 26-34 Years Old
Ordered
  Position:
  4                6275    (4) 35-49 Years Old
Ordered
  Position:
  5                2330    (5) 50-64 Years Old
Ordered
  Position:
  6                1046       (6) 65 or Older
------------------------------------------------
```

The results in *Exhibit 3* show the frequencies for the variable used in the CLASS statement, CATAG6.

Exhibit 4.   XRATIO Output

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: ALCYRTOT <= 365
Estimating the Average Number of Days with at Least One Drink per Week
Done Using XRATIO and XMEAN with a PARAMETER Statement
by: Variable, AGE CATEGORY RECODE (6 LEVELS).

---------------------------------------------------------------------------------------------
Variable
  AGE CATEGORY
    RECODE (6
    LEVELS)               Sample Size     Weighted Size          Estimate        SE Estimate

---------------------------------------------------------------------------------------------
EST1: Using RATIO:
  Days with Drink
  per Week
   Total                     34458        166239053.00            1.69              0.02
   (1) 12-17 Years
     Old                      5143          6445339.59            0.64              0.02
   (2) 18-25 Years
     Old                     15030         26510759.23            1.45              0.01
   (3) 26-34 Years
     Old                      4634         28742534.64            1.61              0.03
   (4) 35-49 Years
     Old                      6275         46702120.81            1.67              0.03
   (5) 50-64 Years
     Old                      2330         37901092.81            1.88              0.05
   (6) 65 or Older            1046         19937205.92            2.13              0.09
---------------------------------------------------------------------------------------------
Data from NSDUH 2010
```

The results in *Exhibit 4* show the output from the XRATIO statement which computes the ratio of ALCYRTOT to 52 which creates an estimate of the average number of days per a week members of the subpopulation have at least one drink of alcohol. The estimate for the total subpopulation, EST1, is 1.69 days/week with a standard error of 0.02. As can be seen, in the subpopulation, the trend with respect towards age is that as the age of the individuals in the age group increases, the more number of days per a week individuals have at least one drink.

Exhibit 5.   XMEAN of ALCYRTOT Output

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: ALCYRTOT <= 365
Estimating the Average Number of Days with at Least One Drink per Week
Done Using XRATIO and XMEAN with a PARAMETER Statement
by: Variable, AGE CATEGORY RECODE (6 LEVELS).


-------------------------------------------------------------------------------------
Variable
  AGE CATEGORY
    RECODE (6
     LEVELS)            Sample Size    Weighted Size        Estimate     SE Estimate
-------------------------------------------------------------------------------------
Mean Drinks per Year
  Total                      34458    166239053.00            87.85            0.99
  (1) 12-17 Years
    Old                       5143      6445339.59            33.46            0.97
  (2) 18-25 Years
    Old                      15030     26510759.23            75.32            0.76
  (3) 26-34 Years
    Old                       4634     28742534.64            83.85            1.67
  (4) 35-49 Years
    Old                       6275     46702120.81            86.94            1.54
  (5) 50-64 Years
    Old                       2330     37901092.81            97.91            2.43
  (6) 65 or Older             1046     19937205.92           110.87            4.57
-------------------------------------------------------------------------------------
Data from NSDUH 2010
```

The results in *Exhibit 5* show the output from the XMEAN statement on ALCYRTOT used to compute the average days of the year members of the subpopulation had at least one or more drinks. As can be seen the estimate for the subpopulation is 87.85 days with a drink(s) per year with a standard error of 0.99.

Exhibit 6.   PARAMETER Statement Computing EST2 Output

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: ALCYRTOT <= 365
Estimating the Average Number of Days with at Least One Drink per Week
Done Using XRATIO and XMEAN with a PARAMETER Statement
by: Variable, AGE CATEGORY RECODE (6 LEVELS).


-------------------------------------------------------------------------------------------
EST2: Using MEAN:
  Days with Drink
  per Week
   Total                        34458      166239053.00           1.69              0.02
    (1) 12-17 Years
     Old                         5143        6445339.59           0.64              0.02
    (2) 18-25 Years
     Old                        15030       26510759.23           1.45              0.01
    (3) 26-34 Years
     Old                         4634       28742534.64           1.61              0.03
    (4) 35-49 Years
     Old                         6275       46702120.81           1.67              0.03
    (5) 50-64 Years
     Old                         2330       37901092.81           1.88              0.05
    (6) 65 or Older              1046       19937205.92           2.13              0.09
-------------------------------------------------------------------------------------------
Data from NSDUH 2010
```

Exhibit 6 displays the results from the PARAMETER statement which utilizes the estimate of days per year with a drink(s) created in the XMEAN statement (*Exhibit 5)* to create an estimate for average days per a week with at least one drink. As can be seen from the results, the estimate for the total subpopulation for days per week with at least one alcoholic drink, EST2, is 1.69 with a standard error of 0.02.

Exhibit 7.   PARAMETER Statement Computing the Difference between EST1 and EST2

```
Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: ALCYRTOT <= 365
Estimating the Average Number of Days with at Least One Drink per Week
Done Using XRATIO and XMEAN with a PARAMETER Statement
by: Variable, AGE CATEGORY RECODE (6 LEVELS).

------------------------------------------------------------------------------------------
Variable
  AGE CATEGORY
    RECODE (6
    LEVELS)              Sample Size    Weighted Size      Estimate     SE Estimate
------------------------------------------------------------------------------------------
Difference between
  EST1 and EST2
   Total                     34458     166239053.00         -0.00            0.00
   (1) 12-17 Years
     Old                      5143       6445339.59          0.00            0.00
   (2) 18-25 Years
     Old                     15030      26510759.23          0.00            0.00
   (3) 26-34 Years
     Old                      4634      28742534.64         -0.00            0.00
   (4) 35-49 Years
     Old                      6275      46702120.81          0.00            0.00
   (5) 50-64 Years
     Old                      2330      37901092.81          0.00            0.00
   (6) 65 or Older            1046      19937205.92          0.00            0.00
------------------------------------------------------------------------------------------
Data from NSDUH 2010
```

The results displayed in *Exhibit 7* display the output from the PARAMETER statement which computes the difference between EST1and EST2 as shown in the output in *Exhibit 4* and *Exhibit 6*, respectively. As can be seen, the difference is zero for both the estimate, and its corresponding standard error across all levels of the CLASS variable, CATAG6, and for the subpopulation total. Thus, EST1 = EST2 and the corresponding standard errors are equal across all levels of CATAG6 and for the total subpopulation. This demonstrates that both methods, using XRATIO, and also XMEAN coupled with a PARAMETER statement are valid ways to compute the same estimate. While the XRATIO statement may have been the most efficient way to generate the estimate, the XMEAN and PARAMETER method yields the same, equally valid result. Both methods can be implemented in the same procedure call.