

VARGEN Example #1

SUDAAN Statements and Results Illustrated

- XPER
- PARAMETER
- SUBPOPX
- CLASS
- DIFFVAR

Input Data Set(s): DA32722P1.SAS7bdat

Example

Using NSDUH 2010 data, this example will estimate the percent of college aged adults that have ever used alcohol or tobacco and compare the results by sex.

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 variables of interest in this example are ALCEVER and CIGEVER. ALCEVER is a numeric representation of whether an individual has ever had a drink of alcohol, where a drink is classified as an entire drink excluding a sip or a taste. CIGEVER is a numeric representation for whether an individual has smoked a cigarette, be it all or part of one. If the individual has ever had a drink then ALCEVER is one. If the individual has ever smoked a cigarette then CIGEVER is one. If the individual has not smoked cigarette then CIGEVER is two. If the individual has not had a drink of alcohol then ALCEVER is two. XPER statements will be used to evaluate the two variables. The resulting variables from the XPER statistics will contain the percent of individuals that have used the specified substance.

After the XPER statistics are compiled then a PARAMETER statement will be used to compute the difference between ALCEVER and CIGEVER.

The CLASS statement is used to separate the output by males and females by using the variable IRSEX. IRSEX is equal to one if the individual is a male and two if female. The CLASS statement is also necessary for the DIFFVAR statement, discussed next, to be used.

The DIFFVAR statement will be used to generate contrasts between males and females for all statistics that are created from the XPER and PARAMETER statistics.

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 the college aged population of the sample. This does not necessarily indicate that the individuals are enrolled in college, but rather of the age of a typical college attendee.

AGE2 is the variable in the NSDUH dataset that indicates age of respondent. If AGE2 equals 7 then the respondent is 18 years old. When AGE2 equals 11 the respondent is between 22 and 23 years of age. We will use the SUBPOPX statement to restrict the analysis to respondents with a value of AGE2 between and including 7 and 11, i.e. the college aged respondents.

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

```
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;
class irsex;
subpopx 7 <= age2 <= 11 / name = "College Aged Adults: Ages 18-23" ;
diffvar irsex=(1 2) / name = "Male vs. Female";
xper cigper : cigever / value = 1 name = "Percent That Have Tried Cigarettes";
xper alcper : alcever/ value = 1 name = "Percent That Have Had a Drink of Alcohol";
parameter alc_cig : alcper - cigper / name = "Percent Tried Alcohol minus Percent Tried Cigs";
setenv colwidth = 15;
rtitle "Difference between Males and Females for trying Cigarettes and Alcohol within the College Aged population";
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  : 14718      Weighted count : 26012245
Denominator degrees of freedom :      60
```

The results displayed in *Exhibit 2* indicate that there are 14,718 sample members in the 18-23 year-old subpopulation as specified by the SUBPOPX statement in *Exhibit 1*. They represent 26,012,245 individuals in the target population.

Exhibit 3. CLASS Variable Frequencies

```

Frequencies and Values for CLASS Variables
by: IMPUTATION REVISED GENDER.

-----
IMPUTATION
REVISED
GENDER          Frequency          Value
-----
Ordered
  Position:
  1              7129             (1) Male
Ordered
  Position:
  2              7589             (2) Female
-----

```

The results from the output displayed in *Exhibit 3* show that in the subpopulation of the 14,718 individuals aged 18-23, 7,129 are male and 7,589 are female.

Exhibit 4. XPER and DIFFVAR of Males and Females for Cigarettes

```

Variance Estimation Method: Taylor Series (WR)
For VARGEN Testing: College Aged Adults: Ages 18-23
Difference between Males and Females for trying Cigarettes and Alcohol within the College Aged
population
by: Variable, SUDAAN Reserved Variable One, Contrast.

for: Variable = Percent That Have Tried Cigarettes.

-----
SUDAAN Reserved
Variable One      Sample Contrast   Weighted          Contrast          SE Contrast
Contrast         Size             Contrast Size     Estimate          Estimate
-----
Total
  Male vs. Female      14718           26012244.92      8.06              1.00
1
  Male vs. Female      14718           26012244.92      8.06              1.00
-----
Data from NSDUH 2010

```

The results from *Exhibit 4* estimate the difference between the percent of males in that subpopulation that have tried cigarettes and females that have. As can be seen from the output the contrast gives us an estimate that 8.06% more males have tried cigarettes than females aged 18-23. Using the standard error contrast estimate of 1.0, the result of contrast shows a statistically significant difference between males and females aged 18-23.

Exhibit 5. XPER and DIFFVAR of Males and Females for Alcohol

Variance Estimation Method: Taylor Series (WR)
 For VARGEN Testing: College Aged Adults: Ages 18-23
 Difference between Males and Females for trying Cigarettes and Alcohol within the College Aged population
 by: Variable, SUDAAN Reserved Variable One, Contrast.
 for: Variable = Percent That Have Had a Drink of Alcohol.

SUDAAN Reserved Variable One Contrast	Sample Contrast Size	Weighted Contrast Size	Contrast Estimate	SE Contrast Estimate
Total				
Male vs. Female	14718	26012244.92	3.45	0.82
1				
Male vs. Female	14718	26012244.92	3.45	0.82

Data from NSDUH 2010

The results from *Exhibit 5* show that the estimated difference between males in the subpopulation that have had a drink of alcohol and females is 3.45%. The standard error for the contrast estimate is 0.82. This standard error gives a statistically significant difference between males and females for the percent of individual that have had a drink of alcohol in the subpopulation of individuals aged 18-23.

Exhibit 6. PARAMETER and DIFFVAR of Males and Females

Variance Estimation Method: Taylor Series (WR)
 For VARGEN Testing: College Aged Adults: Ages 18-23
 Difference between Males and Females for trying Cigarettes and Alcohol within the College Aged population
 by: Variable, SUDAAN Reserved Variable One, Contrast.
 for: Variable = Percent Tried Alcohol minus Percent Tried Cigs.

SUDAAN Reserved Variable One Contrast	Sample Contrast Size	Weighted Contrast Size	Contrast Estimate	SE Contrast Estimate
Total				
Male vs. Female	14718	26012244.92	-4.61	1.07
1				
Male vs. Female	14718	26012244.92	-4.61	1.07

Data from NSDUH 2010

The results from *Exhibit 6* show the output from the PARAMETER statement which takes the difference between the variables that were created through the XPER statements. The difference between ALCPER (the percent of individuals who have tried alcohol) and CIGPER (the percent of individuals who have tried cigarettes) is compiled separately for both males and females. The resulting variable is alc_cig. The output then shows the result of the contrast which shows the difference of alc_cig between males and females. The estimate for this is -4.61%. This can be interpreted as the difference between percent of individuals who have tried alcohol compared to cigarettes is estimated to be 4.61% less for males than females ages 18-23. The standard error for the contrast estimate is 1.07, which yields a statistically significant contrast estimate.