CROSSTAB Example #4

SUDAAN Statements and Results Illustrated

- TEST
- PRINT STEST option
- SUBPOPX
- SETENV
- RFORMAT

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Among adults with arthritis, estimate the type of arthritis, by gender, using NHANES III.

Solution

The data set is adults aged 17 and older from NHANES III. All variables in this example are from the home interview component of NHANES III, and all six years of data are analyzed. Thus, the sample weight variable is WTPFQX6, and the stratification and PSU variables are SDPSTRA6 and SDPPSU6, respectively. The SAS-Callable code for this example is presented in ***Exhibit 1***.

The SUBPOPX statement is used to subset the data set to adults with arthritis, because these are the only subjects who were asked the question about type of arthritis (HAC1B). In addition to the two types (osteoarthritis and rheumatoid arthritis), several subjects replied that they did not know what type of arthritis they had. These subjects are included in the analysis since there are too many of them to exclude.

The TABLES statement requests a cross-tabulation of type of arthritis (row variable) by gender (column variable); hence, column percent is requested. The TEST statement will produce a Pearson-type hypothesis test of general association between arthritis type and gender. All test statistics are requested for the CHISQ hypothesis of general association (the CHISQ hypothesis test statistics are in the STEST output group).

The SETENV statement manipulates the printout so that all columns of the table can fit on a single page. The PRINT statement tailors the output to get specific statistics in a particular format.

This example was run in SAS-Callable SUDAAN, and the SAS program and *.LST files are provided.

Exhibit 1. SAS-Callable SUDAAN Code

```
libname in v604 "c:\10winbetatest\examplemanual\crosstab";

options pagesize=70 linesize=80;
proc format;
  value sex 1="1=Male"
            2="2=Female";
  value type 1="1=Rheumatoid"
             2="2=Osteo"
             3="3=Don't Know";

PROC CROSSTAB DATA=in.HANES3S3 FILETYPE=SAS DESIGN=WR DEFT1;
  NEST SDPSTRA6 SDPPSU6;
  WEIGHT WTPFQX6;

  SUBPOPX HAC1A=1 / NAME="TOLD BY MD HAVE ARTHRITIS";

  CLASS HSSEX HAC1B;
  TABLES HAC1B*HSSEX;  /* HAC1B IS TYPE ARTHRITIS */
  TEST CHISQ / all;

  SETENV ROWWIDTH=15 COLWIDTH=10 LABWIDTH=27;
  PRINT NSUM="SAMSIZE" WSUM="POPSIZE" COLPER SECOL DEFFCOL="DEFF1COL" /
        STEST=default NSUMFMT=F9.0 WSUMFMT=F9.0 STESTVALFMT=F10.2 SPVALFMT=F8.4
        SDFFMT=F8.0 SADJDFFMT=F8.0;
  rformat hssex sex.;
  rformat hac1b type.;
  RTITLE "TYPE OF ARTHRITIS, BY SEX, AMONG THOSE WITH ARTHRITIS";
  RFOOTNOTE "NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, ADULTS (17+)";
```

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


DESIGN SUMMARY: Variances will be computed using the Taylor Linearization
Method, Assuming a With Replacement (WR) Design
    Sample Weight: WTPFQX6
    Stratification Variables(s): SDPSTRA6
    Primary Sampling Unit: SDPPSU6


Number of observations read    :  20050    Weighted count :187647206
Observations in subpopulation  :   4298    Weighted count : 32666641
Denominator degrees of freedom :     49
```

The SUBPOPX statement restricts the analysis to adults who were told by a doctor that they had arthritis. SUDAAN identified 4,298 sample adults in the subpopulation (***Exhibit 2***), and they represent an estimated 32,666,641 adults in the population with arthritis.

Next, SUDAAN displays the frequencies of the CLASS variables (***Exhibit 3***).

Exhibit 3. CLASS Variable Frequencies

```
Frequencies and Values for CLASS Variables

by: Sex.
-------------------------------------
Sex             Frequency       Value
-------------------------------------
Ordered
  Position:
  1                  1570      1=Male
Ordered
  Position:
  2                  2728    2=Female
-------------------------------------
```

Exhibit 3. CLASS Variable Frequencies-cont.

```
Frequencies and Values for CLASS Variables

by: Type arthritis:rheumatoid,osteoarthritis.
-----------------------------------------------
Type Arthritis:
  rheumatoid
  osteoarthritis     Frequency           Value
-----------------------------------------------
Ordered
  Position:
  1                       840    1=Rheumatoid
Ordered
  Position:
  2                       825         2=Osteo
Ordered
  Position:
  3                      2574    3=Don't Know
-----------------------------------------------
```

SUDAAN then displays the results from the PRINT statement (***Exhibit 4***):

Exhibit 4. HAC1B*HSSEX Crosstabulation

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: TOLD BY MD HAVE ARTHRITIS

TYPE OF ARTHRITIS, BY SEX, AMONG THOSE WITH ARTHRITIS

by: Type arthritis:rheumatoid,osteoarthritis, Sex.

Type arthritis:rheumatoid,osteoarthritis		Sex		
		Total	1=Male	2=Female
Total	SAMSIZE	4239	1546	2693
	POPSIZE	32130419	11567703	20562716
	Col Percent	100.00	100.00	100.00
	SE Col Percent	0.00	0.00	0.00
	DEFF1COL	.	.	.
1=Rheumatoid	SAMSIZE	840	278	562
	POPSIZE	6441250	1992084	4449166
	Col Percent	20.05	17.22	21.64
	SE Col Percent	1.03	1.49	1.25
	DEFF1COL	2.28	1.92	2.04
2=Osteo	SAMSIZE	825	241	584
	POPSIZE	7736008	2537690	5198317
	Col Percent	24.08	21.94	25.28
	SE Col Percent	1.44	1.80	1.71
	DEFF1COL	3.92	2.34	3.41
3=Don't Know	SAMSIZE	2574	1027	1547
	POPSIZE	17953162	7037928	10915233
	Col Percent	55.88	60.84	53.08
	SE Col Percent	1.67	2.24	1.81
	DEFF1COL	3.90	2.61	2.88

NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, ADULTS (17+)

Among adults with arthritis, an estimated 56% do not know what type of arthritis they have; while 20% identify rheumatoid arthritis and 24% identify osteoarthritis (***Exhibit 4***). Note that the three percentages on the Total row add to 100%, as do the three percentages on each of the Male and Female rows.

The gender-specific estimated percentages for "don't know" are 61% for males and 53% for females.

The comparison sampling plan for the requested DEFT1 is a simple random sample of 4,239 adults from all adults with arthritis. The design effects are not as large here as when estimating the prevalence of arthritis for the entire population and using DEFT1 (see ***Example 5***), most likely because the subpopulation analysis here has a smaller average number of elements per PSU (cluster).

Exhibit 5. Stratum-Specific Tests of Hypotheses for HAC1B*HSSEX

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: TOLD BY MD HAVE ARTHRITIS

TYPE OF ARTHRITIS, BY SEX, AMONG THOSE WITH ARTHRITIS

Test Statistics for Stratum-Specific Hypotheses
Variable HAC1B by Variable HSSEX

Hypothesis Test Test Statistic	DF	Adj DF	Test Value	P-Value
CHISQ (Obs - Exp)				
Wald chi-square	2	.	12.62	0.0018
Wald-F	2	.	6.31	0.0036
Adj Wald F	2	.	6.18	0.0041
Satterthwaite-adj chi-sq	2	2	11.48	0.0032
Satterthwaite-adj F	2	2	5.78	0.0057

NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, ADULTS (17+)

The 5 test statistics (***Exhibit 5***) used to evaluate the CHISQ hypothesis of no association indicates that males and females with arthritis differ significantly on the type of arthritis reported. The difference seems to be that females are less likely to say "don't know."