

CROSSTAB Example #6

SUDAAN Statements and Results Illustrated

- Small Percentage Confidence Interval (SPCI)
- ROWPER, ROWSPCI
- STYLE option
- RFORMAT
- SUBPOPX

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Using the NHANES III data, gives an example of the small percentage confidence interval option.

Solution

This example illustrates the small percentage confidence interval (SPCI) enhancement in CROSSTAB. The data set consists of a subset of adults aged 17 and older from *NHANES III*, a cross-sectional sample survey of the civilian, non-institutionalized population aged 2 months or older, fielded during 1988-1994. 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 used in this example is displayed in *Exhibit 1*.

This example estimates the percentage of adults aged 17-22 years that have been told by a doctor that they have arthritis. These estimates are computed by year of age. Since all ages in the subpopulation will likely have a very low prevalence of arthritis, we specify SMCONF=5.0 on the PROC statement to trigger the SPCI method for all row percentages less than or equal to 5% (see *Exhibit 1*). In the 2-way table of age-by-arthritis (HSAGEIR*HAC1A), row percentages tell us the percentage of arthritis cases in each age group. Therefore, we request row percentages (ROWPER) on the PRINT statement. The requested ROWSPCI will serve as a flag in the PRINT tables to indicate when the SPCI was used.

The SUBPOPX statement is used to subset the population to those between less than 22 years of age; the CLASS statement defines the categorical variables of interest; and the TABLES statement defines the rows and columns of the cross-classification of interest.

The PRINT statement is used in this example to request only the statistics of interest, to obtain the results in a list format (STYLE=NCHS) instead of the default box format, and to customize the formats of certain printed statistics.

The RFORMAT statements associate a SAS format with one of the variables used in the CROSSTAB procedure.

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 "c:\10winbetatest\examplemanual\crosstab";

proc format;
  value yn 1="Yes"
          2="No";

proc crosstab data=in.hanes3s5 design=wr SMCNF=5.0;
  nest sdpstra6 sdpps6;
  weight wtpfqx6;

  subpopx HSAGEIR < 22; /* 17-21 year-olds */
  class hsageir HAC1A;
  table hsageir*HAC1A; /* age-by-arthritis */

  print nsum rowper lowrow uprow ROWSPCI /
        rowperfmt=f7.2 lowrowfmt=f6.2 uprowfmt=f6.2 nsumfmt=f6.0 style=nchs;
  RFORMAT HAC1A yn.;
  RTITLE "SPCI Example Using NHANES III Data: Arthritis Prevalence in 17-21 Year-Olds";
```

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   : 1967       Weighted count : 17314389
Denominator degrees of freedom : 49
```

The results displayed in *Exhibit 2* indicate that there are 1,967 sample members in the 17-22 year-old subpopulation. They represent 17,314,389 individuals in the target population.

The frequency distribution for each CLASS variable is presented next (*Exhibit 3*). Of the 1,967 sample members, only 29 report having been told by doctor that they had arthritis.

Exhibit 3. CLASS Variable Frequencies

Frequencies and Values for CLASS Variables

by: Age at interview (screener) - qty.

Age at interview (screener) - qty	Frequency	Value

Ordered Position: 1	432	17
Ordered Position: 2	403	18
Ordered Position: 3	390	19
Ordered Position: 4	365	20
Ordered Position: 5	377	21

Exhibit 3. CLASS Variable Frequencies-cont.

Frequencies and Values for CLASS Variables

by: Doctor ever told you had: arthritis.

Doctor ever told you had: arthritis	Frequency	Value

Ordered Position: 1	29	Yes
Ordered Position: 2	1938	No

Exhibit 4. HSAGEIR*HAC1A Crosstabulation

```

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: HSAGEIR < 22

SPCI Example Using NHANES III Data: Arthritis Prevalence in 17-21 Year-Olds

by: Age at interview (screener) - qty, Doctor ever told you had: arthritis.
-----
Age at interview (screener) - qty Doctor ever told you had: arthritis
Sample Size Row Percent Lower 95% Limit ROWPER Upper 95% Limit ROWPER
-----
Total
  Total 1967 100.00 . .
  Yes 29 1.64 0.90 2.74 ^
  No 1938 98.36 97.26 99.10 ^
17
  Total 432 100.00 . .
  Yes 9 2.15 0.86 4.40 ^
  No 423 97.85 95.60 99.14 ^
18
  Total 403 100.00 . .
  Yes 2 0.55 0.02 2.71 ^
  No 401 99.45 97.29 99.98 ^
19
  Total 390 100.00 . .
  Yes 3 1.77 0.20 6.40 ^
  No 387 98.23 93.60 99.80 ^
20
  Total 365 100.00 . .
  Yes 10 2.14 0.51 5.77 ^
  No 355 97.86 94.23 99.49 ^
21
  Total 377 100.00 . .
  Yes 5 1.58 0.09 7.04 ^
  No 372 98.42 92.96 99.91 ^
-----
LOWROW and UPROW use small percentage confidence interval calculation when ROWPER <= 5.00% or
ROWPER >= 95.00%

```

The results in *Exhibit 4* indicate for each year of age and overall, the percentage of people reporting arthritis (row percentage in this age*arthritis table) is always $\leq 5\%$ (yes) or $\geq 95\%$ (no). *SMCONF=5.0* is the criteria specified on the PROC statement to trigger the SPCI, and therefore all confidence limits for the row percentages in this example are based on the SPCI approach. The last column in the table (from ROWSPCI keyword on PRINT statement) indicates when the SPCI was used. CROSSTAB also adds the table footnote to explain the meaning of the ^ symbol.