

CROSSTAB Example #8

SUDAAN Statements and Results Illustrated

- Stratum-specific Chi-square (CHISQ) Test
- Stratum-adjusted Cochran-Mantel-Haenszel (CMH) Test
- ANOVA-type (ACMH) Test
- ALL Test option
- DISPLAY option

Input Data Set(s): NHANES3S3.SAS7bdat

Example

This example illustrates the variety of hypotheses and test statistics now available on the TEST statement in CROSSTAB.

Solution

The data set consists 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 SUDAAN code used for this example is displayed in *Exhibit 1*.

This example uses the *NHANES III* dataset to evaluate whether having been diagnosed with arthritis (HAC1A, 1=yes, 2=no) is significantly associated with the self-reported condition of someone's natural teeth (TEETHSTAT, collapsed from HAQ1 as excellent/very good, good, fair, or poor/no natural teeth). We also want to evaluate this hypothesis after adjusting for age group (17-49 vs. 50+) and self-reported general health status (HLTHSTAT, collapsed from HAB1 as excellent/very good, good, or fair/poor).

In the SAS-Callable SUDAAN code below (*Exhibit 1*), the following statements are used:

- TABLES statement requests: 1) a single 2-way table of arthritis-by-teeth status, and 2) a stratified (by age and health status) two-way table of arthritis-by-teeth status. All of these variables are listed on the CLASS statement.
- TEST statement requests the *stratum-specific* CHISQ test of general association for each table request—based on (observed – expected), analogous to Pearson chi-square for non-survey data—plus the *stratum-adjusted* Cochran-Mantel-Haenszel tests of general association (CMH), trend (TCMH), and the ANOVA-type test (ACMH) for each table request. There is an option to request all five available test statistics for each hypothesis (ALL), as well as to display the scores used for TCMH and ACMH hypotheses (DISPLAY).
- PRINT statement requests only the statistics of interest: sample and estimated population size, column percentages, and standard errors for each TABLES request, as well as the defaults from the STEST group (statistics for CHISQ hypothesis) and ATEST group (statistics for CMH, TCMH, and

ACMH hypotheses). It also changes statistic labels and formats to customize the printed results (SETENV helps with this also).

- The RFORMAT statements associate SAS formats with the CLASS variables.

This example was run in SAS-Callable SUDAAN, and the SAS program and *.LST files are provided. The SAS programming statements come first, and they are used to create the variables of interest: HAC1A (9's recoded to missing); HLTHSTAT created from HAB1; TEETHSTAT created from HAQ1; and AGEGRP2 created from HSAGEIR.

Exhibit 1. SAS-Callable SUDAAN Code

```
libname in "c:\10winbetatest\examplemanual\crosstab";
options nocenter linesize=85 pagesize=60;

proc format;
  value age2fmt 1="1=17-49"
              2="2=50+";
  Value yesnofmt 1="1=Yes"
                2="2=No";
  value hlthfmt 1="1=Excellent/Very Good"
               2="2=Good"
               3="3=Fair/Poor";
  value teethfmt 1="1=Excellent/Very Good"
                 2="2=Good"
                 3="3=Fair"
                 4="4=Poor/No Natural Teeth";

data one; set in.nhanes3;
  if hacla=9 then hacla=.;

  if hab1 in (8,9) then hab1=.;
  else if hab1 in (1,2) then hlthstat=1;
  else if hab1 in (3) then hlthstat=2;
  else if hab1 in (4,5) then hlthstat=3;

  if haq1 in (88,99) then haq1=.;
  else if haq1 in (1,2) then teethstat=1;
  else if haq1 in (3) then teethstat=2;
  else if haq1 in (4) then teethstat=3;
  else if haq1 in (5,6) then teethstat=4;

  if 17 le hsageir le 49 then agegrp2=1;
  else if hsageir ge 50 then agegrp2=2;

  label agegrp2="Age Group";

PROC SORT data=one; by sdpstra6 sdpps6;

PROC CROSSTAB DATA=one FILETYPE=SAS DESIGN=WR deft1;
  NEST SDPSTRA6 SDPPSU6;
  WEIGHT WTPFQX6;

  CLASS HAC1A AGEGRP2 HLTHSTAT TEETHSTAT;
  TABLES HAC1A*TEETHSTAT AGEGRP2*HLTHSTAT*HAC1A*TEETHSTAT;
  TEST CHISQ CMH ACMH TCMH / display ALL;

  SETENV ROWWIDTH=9 LBLWIDTH=8 COLWIDTH=8 LABWIDTH=27;
  rformat agegrp2 age2fmt.;
  rformat hacla yesnofmt.;
  rformat hlthstat hlthfmt.;
  rformat teethstat teethfmt.;
  PRINT nsum="SAMSIZE" wsum="POPSIZE" colper="COL %" secol="SE" /
        STEST=default ATEST=default spvalfmt=f8.4 apvalfmt=f8.4
        stestvalfmt=f6.2 atestvalfmt=f6.2 colperfmt=f8.2 secolfmt=f8.2
        nsumfmt=f6.0 wsumfmt=f9.0;
  RTITLE "NHANES 3: TEST Statement Hypotheses for Stratified and Non-Stratified
  Tables";
```

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
Denominator degrees of freedom :      49
```

There are 20,050 sample members in this NHANES-III dataset (*Exhibit 2*).

Below (*Exhibit 3*) are the unweighted frequency distributions for the variables on the CLASS statement (Arthritis, Age Group, Health Status, Teeth Status).

Exhibit 3. CLASS Variable Frequencies

```

Frequencies and Values for CLASS Variables

by: Doctor ever told you had: arthritis.
-----
Doctor ever
told you
had:
arthritis      Frequency      Value
-----
Ordered
Position:
1              4298      1=Yes
Ordered
Position:
2              15748     2=No
-----
```

Exhibit 3. CLASS Variable Frequencies-cont.

```

Frequencies and Values for CLASS Variables

by: Age Group.
-----
Age Group      Frequency      Value
-----
Ordered
Position:
1              11396     1=17-49
Ordered
Position:
2              8654      2=50+
-----
```

Exhibit 3. CLASS Variable Frequencies-cont.

```
Frequencies and Values for CLASS Variables
by: HLTHSTAT.
-----
HLTHSTAT      Frequency      Value
-----
Ordered
  Position:
    1          7813      1=Excellent/Very Good
Ordered
  Position:
    2          7191          2=Good
Ordered
  Position:
    3          5033          3=Fair/Poor
-----
```

Exhibit 3. CLASS Variable Frequencies-cont.

```
Frequencies and Values for CLASS Variables
by: TEETHSTAT.
-----
TEETHSTAT      Frequency      Value
-----
Ordered
  Position:
    1          3881      1=Excellent/Very Good
Ordered
  Position:
    2          5666          2=Good
Ordered
  Position:
    3          4891          3=Fair
Ordered
  Position:
    4          5595      4=Poor/No Natural Teeth
-----
```

Below (*Exhibit 4*) is the display of scores assigned to the teeth status and arthritis variables for the TCMH and ACMH hypotheses. Since there is no SCORES statement, all scores are defined by the values of the variables on the CLASS statement.

Exhibit 4. Scores for TCMH and ACMH Hypotheses

Scores Used in Computation of ACMH (column only), TCMH (row and column)
 by: Doctor ever told you had: arthritis.

```

-----
Doctor ever
  told you
  had:
  arthritis      Value      Score
-----
Ordered
  Position:
  1              1=Yes      1
Ordered
  Position:
  2              2=No      2
-----
  
```

Exhibit 4. Scores for TCMH and ACMH Hypotheses-cont.

Scores Used in Computation of ACMH (column only), TCMH (row and column)
 by: TEETHSTAT.

```

-----
TEETHSTAT              Value      Score
-----
Ordered
  Position:
  1              1=Excellent/Very Good      1
Ordered
  Position:
  2              2=Good      2
Ordered
  Position:
  3              3=Fair      3
Ordered
  Position:
  4              4=Poor/No Natural Teeth      4
-----
  
```

Next are various statistics from the first TABLES request (*Exhibit 5*): arthritis-by-teeth status (HAC1A*TEETHSTAT). This table contains column percentages and related statistics, as requested on the PRINT statement (NSUM WSUM COLPER SECOL). In the interest of brevity, the stratified version of this arthritis-by-teeth status table (cross-classified by levels of age group and health status)—as generated by the TABLES request AGEGRP2*HLTHSTAT*HAC1A*TEETHSTAT—is not presented here, but was included in the PRINT tables.

Overall, the prevalence of diagnosed arthritis in the adult population increases from 11.2% for those with self-reported Excellent/Very Good Teeth, steadily up to 31.9% for those with Poor/No Natural Teeth.

Exhibit 5. HAC1A*TEETHSTAT Crosstabulation

Variance Estimation Method: Taylor Series (WR)

NHANES 3: TEST Statement Hypotheses for Stratified and Non-Stratified Tables

by: Doctor ever told you had: arthritis, TEETHSTAT.

		TEETHSTAT				
		Total	1=Excell- ent/Very Good	2=Good	3=Fair	4=Poor/No Natural Teeth
Total	SAMSIZE	20029	3881	5664	4890	5594
	POPSIZE	187505214	50781274	60305634	37175914	39242392
	COL %	100.00	100.00	100.00	100.00	100.00
	SE	0.00	0.00	0.00	0.00	0.00
1=Yes	SAMSIZE	4291	518	899	882	1992
	POPSIZE	32625665	5703311	7964943	6442465	12514945
	COL %	17.40	11.23	13.21	17.33	31.89
	SE	0.51	0.85	0.68	0.93	0.93
2=No	SAMSIZE	15738	3363	4765	4008	3602
	POPSIZE	154879549	45077963	52340690	30733449	26727447
	COL %	82.60	88.77	86.79	82.67	68.11
	SE	0.51	0.85	0.68	0.93	0.93

Below (*Exhibit 6*) are the results from the CHISQ hypothesis of general association between arthritis and teeth status (not stratified) from the TABLES request HAC1A*TEETHSTAT. Using each of the 5 test statistics, there is a strong association.

In the interest of brevity, the CHISQ hypothesis tested within each level of age group by general health status (generated by the TABLES request AGEGRP2*HLTHSTAT*HAC1A*TEETHSTAT) is not presented here, but was included in the PRINT tables.

Exhibit 6. Stratum-Specific Hypothesis Tests for HAC1A*TEETHSTAT

Variance Estimation Method: Taylor Series (WR)

NHANES 3: TEST Statement Hypotheses for Stratified and Non-Stratified Tables

Test Statistics for Stratum-Specific Hypotheses
Variable HAC1A by Variable TEETHSTAT

Hypothesis Test	Test Statistic	DF	Adj DF	Test Value	P-Value
CHISQ (Obs - Exp)					
	Wald chi-square	3	.	238.56	0.0000
	Wald-F	3	.	79.52	0.0000
	Adj Wald F	3	.	76.28	0.0000
	Satterthwaite-adj chi-sq	3	2.55	209.73	0.0000
	Satterthwaite-adj F	3	2.55	82.10	0.0000

Below (*Exhibit 7*) are the results from the CMH, TCMH, and ACMH hypotheses for the 2-way table of arthritis by teeth status, corresponding to the TABLES request HAC1A*TEETHSTAT. Since there is no

stratification in this 2-way table, the CMH general association hypothesis is equivalent to the CHISQ hypothesis presented on the previous page. However, we now also get the test for trend and ANOVA-type test for this table. There is a significant trend in arthritis prevalence across teeth status. The ANOVA-type test (ACMH) is equivalent to the trend test (TCMH) when the row variable has only 2 levels.

Exhibit 7. Stratum-Adjusted Hypothesis Tests for HAC1A*TEETHSTAT

```

Variance Estimation Method: Taylor Series (WR)

NHANES 3: TEST Statement Hypotheses for Stratified and Non-Stratified Tables

Test Statistics for Stratum-Adjusted Hypotheses
Variable HAC1A by Variable TEETHSTAT

-----
Hypothesis Test
Test Statistic
-----
DF      Adj DF  Test Value  P-Value
-----
CMH General Association
Wald chi-square      3      .      238.56      0.0000
Wald-F               3      .       79.52      0.0000
Adj Wald F          3      .       76.28      0.0000
Satterthwaite-adj chi-sq  3      2.55    209.73      0.0000
Satterthwaite-adj F    3      2.55     82.10      0.0000
CMH Trend
Wald chi-square      1      .      194.85      0.0000
Wald-F               1      .      194.85      0.0000
Adj Wald F           1      .      194.85      0.0000
Satterthwaite-adj chi-sq  1      1.00    194.85      0.0000
Satterthwaite-adj F    1      1.00    194.85      0.0000
CMH ANOVA (Row Means)
Wald chi-square      1      .      194.85      0.0000
Wald-F               1      .      194.85      0.0000
Adj Wald F           1      .      194.85      0.0000
Satterthwaite-adj chi-sq  1      1.00    194.85      0.0000
Satterthwaite-adj F    1      1.00    194.85      0.0000
-----

```

Below (*Exhibit 8*) are the results from the CMH, TCMH, and ACMH hypotheses for the *stratified* 2-way table of arthritis by teeth status, generated by the following TABLES request:

AGEGRP2*HLTHSTAT*HAC1A*TEETHSTAT.

After adjustment for age group and general health status, there is still a significant general association and trend in arthritis prevalence across teeth status (although not nearly as significant after adjustment). The ANOVA-type test is again equivalent to the trend test for tables, where the row variable has only 2 levels.

Exhibit 8. Stratum-Adjusted Hypothesis Tests for HAC1A*TEETHSTAT, Controlling for AGEGRP2*HLTHSTAT

Variance Estimation Method: Taylor Series (WR)

NHANES 3: TEST Statement Hypotheses for Stratified and Non-Stratified Tables

Test Statistics for Stratum-Adjusted Hypotheses
 Variable HAC1A by Variable TEETHSTAT
 Controlling for: Variable AGEGRP2 and HLTHSTAT

Hypothesis Test				
Test Statistic	DF	Adj DF	Test Value	P-Value

CMH General Association				
Wald chi-square	3	.	17.47	0.0006
Wald-F	3	.	5.82	0.0017
Adj Wald F	3	.	5.59	0.0023
Satterthwaite-adj chi-sq	3	2.58	10.52	0.0098
Satterthwaite-adj F	3	2.58	4.08	0.0152
CMH Trend				
Wald chi-square	1	.	6.27	0.0123
Wald-F	1	.	6.27	0.0156
Adj Wald F	1	.	6.27	0.0156
Satterthwaite-adj chi-sq	1	1.00	6.27	0.0123
Satterthwaite-adj F	1	1.00	6.27	0.0156
CMH ANOVA (Row Means)				
Wald chi-square	1	.	6.27	0.0123
Wald-F	1	.	6.27	0.0156
Adj Wald F	1	.	6.27	0.0156
Satterthwaite-adj chi-sq	1	1.00	6.27	0.0123
Satterthwaite-adj F	1	1.00	6.27	0.0156
