CROSSTAB Example #2

SUDAAN Statements and Results Illustrated

- SETENV optional statement
- CMH test (Cochran-Mantel-Haenszel)
- PRINT
- RFORMAT
- SEWGT

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Using NHANES III adult data, determine whether gender is related to arthritis, after controlling for age.

Solution

The data set comprises 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.

This example was run in SAS-Callable SUDAAN, and the SAS program and *.LST files are provided. The TABLES statement requests a two-way table of gender (row variable HSSEX) by arthritis (column variable HAC1A), for each level of age group (AGEGRP4) (see ***Exhibit 1***). The TEST statement will produce a Pearson-type chi-squared test of arthritis with gender for each level of age group. In addition, the TEST statement requests a Cochran-Mantel-Haenszel test for general association, a Pearson type chi-square test of arthritis with gender that summarizes the "observed – expected" deviations over age groups (or over age "strata," as most epidemiologists would say). Recall that, in the context of NHANES III data, "Observed" and "Expected" are population estimates, not sample sizes.

Exhibit 1. SAS-Callable SUDAAN Code

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

options pagesize=70 linesize=80;
proc format;
  value yesno 1="1=Yes"
              2="2=No";
  value age 1="17-34"
            2="35-49"
            3="50-64"
            4="65-90+";
  value sex 1="1=Male"
            2="2=Female";


PROC CROSSTAB DATA=in.hanes3s3 FILETYPE=SAS DESIGN=WR;
  NEST SDPSTRA6 SDPPSU6;
  WEIGHT WTPFQX6;

  CLASS AGEGRP4 HSSEX HAC1A;
  TABLES AGEGRP4*HSSEX*HAC1A;
  TEST CHISQ CMH;

  SETENV ROWWIDTH=12 COLWIDTH=10 LABWIDTH=25;
  PRINT  NSUM="SAMSIZE" WSUM="POPSIZE" ROWPER SEROW / STEST=default ATEST=default
         WSUMFMT=F9.0 SEROWFMT=F7.3 STESTVALFMT=F10.2 SDFFMT=F9.0 ADFFMT=F9.0
         ATESTVALFMT=F10.2 ;
  rformat agegrp4 age.;
  rformat hac1a yesno.;
  rformat hssex sex.;
  RTITLE "Association Between GENDER and ARTHRITIS, Controlling for AGE
          (ADULTS 17+)"
  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
Denominator degrees of freedom :     49
```

The SAS *.LST file in ***Exhibit 2*** shows that SUDAAN read in 20,050 adults in the data set. The value of the sampling weight variable WTPFQX6, summed over these 20,050 adults, is 187,647,206, an estimate of the average U.S. adult (aged 17+) civilian, non-institutionalized population during 1988-1994. The denominator degrees of freedom (ddf) for NHANES III is calculated by SUDAAN by its identification of 98 "pseudo-PSUs" and 49 "pseudo-strata" in the data set (*i.e.*, 49 ddf = 98 PSUs – 49 strata).

Next, SUDAAN displays the frequencies of the CLASS variables (not included here, see previous example).

SUDAAN then displays the results from the PRINT statement (***Exhibit 3***). Following are five two-way tables of gender by arthritis—one for each level of age group and one for all ages combined (labelled

AGEGRP4=Total in the output). The estimated prevalence of arthritis is higher in females than in males at each of the four levels of AGEGRP4.

Exhibit 3. AGEGRP4*HSSEX*HAC1A Crosstabulation

Variance Estimation Method: Taylor Series (WR)
Association Between GENDER and ARTHRITIS, Controlling for AGE (ADULTS 17+)

by: AGEGRP4, Sex, Doctor ever told you had: arthritis.
for: AGEGRP4 = Total.

Sex		Doctor ever told you had: arthritis		
		Total	1=Yes	2=No
Total	SAMSIZE	20046	4298	15748
	POPSIZE	187611487	32666641	154944847
	Row Percent	100.00	17.41	82.59
	SE Row Percent	0.000	0.510	0.510
1=Male	SAMSIZE	9399	1570	7829
	POPSIZE	89630819	11789474	77841345
	Row Percent	100.00	13.15	86.85
	SE Row Percent	0.000	0.640	0.640
2=Female	SAMSIZE	10647	2728	7919
	POPSIZE	97980668	20877167	77103501
	Row Percent	100.00	21.31	78.69
	SE Row Percent	0.000	0.591	0.591

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

Exhibit 3. AGEGRP4*HSSEX*HAC1A Crosstabulation-cont.

Variance Estimation Method: Taylor Series (WR)
Association Between GENDER and ARTHRITIS, Controlling for AGE (ADULTS 17+)

by: AGEGRP4, Sex, Doctor ever told you had: arthritis.
for: AGEGRP4 = 17-34.

Sex		Doctor ever told you had: arthritis		
		Total	1=Yes	2=No
Total	SAMSIZE	6900	228	6672
	POPSIZE	71857480	2822848	69034632
	Row Percent	100.00	3.93	96.07
	SE Row Percent	0.000	0.409	0.409
1=Male	SAMSIZE	3262	76	3186
	POPSIZE	35922564	1184692	34737872
	Row Percent	100.00	3.30	96.70
	SE Row Percent	0.000	0.622	0.622
2=Female	SAMSIZE	3638	152	3486
	POPSIZE	35934916	1638156	34296760
	Row Percent	100.00	4.56	95.44
	SE Row Percent	0.000	0.532	0.532

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

Exhibit 3. AGEGRP4*HSSEX*HAC1A Crosstabulation-cont.

Variance Estimation Method: Taylor Series (WR)
Association Between GENDER and ARTHRITIS, Controlling for AGE (ADULTS 17+)

by: AGEGRP4, Sex, Doctor ever told you had: arthritis.
for: AGEGRP4 = 35-49.

Sex		Doctor ever told you had: arthritis		
		Total	1=Yes	2=No
Total	SAMSIZE	4496	557	3939
	POPSIZE	53642570	6647246	46995324
	Row Percent	100.00	12.39	87.61
	SE Row Percent	0.000	0.650	0.650
1=Male	SAMSIZE	2069	180	1889
	POPSIZE	25920111	2706268	23213843
	Row Percent	100.00	10.44	89.56
	SE Row Percent	0.000	0.985	0.985
2=Female	SAMSIZE	2427	377	2050
	POPSIZE	27722459	3940978	23781481
	Row Percent	100.00	14.22	85.78
	SE Row Percent	0.000	0.936	0.936

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

Exhibit 3. AGEGRP4*HSSEX*HAC1A Crosstabulation-cont.

Variance Estimation Method: Taylor Series (WR)
Association Between GENDER and ARTHRITIS, Controlling for AGE (ADULTS 17+)

by: AGEGRP4, Sex, Doctor ever told you had: arthritis.
for: AGEGRP4 = 50-64.

Sex		Doctor ever told you had: arthritis		
		Total	1=Yes	2=No
Total	SAMSIZE	3401	1072	2329
	POPSIZE	32114722	9555128	22559594
	Row Percent	100.00	29.75	70.25
	SE Row Percent	0.000	0.892	0.892
1=Male	SAMSIZE	1625	380	1245
	POPSIZE	15156961	3238161	11918800
	Row Percent	100.00	21.36	78.64
	SE Row Percent	0.000	1.412	1.412
2=Female	SAMSIZE	1776	692	1084
	POPSIZE	16957761	6316967	10640794
	Row Percent	100.00	37.25	62.75
	SE Row Percent	0.000	1.479	1.479

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

Exhibit 3. AGEGRP4*HSSEX*HAC1A Crosstabulation-cont.

Variance Estimation Method: Taylor Series (WR)
Association Between GENDER and ARTHRITIS, Controlling for AGE (ADULTS 17+)

by: AGEGRP4, Sex, Doctor ever told you had: arthritis.
for: AGEGRP4 = 65-90+.

Sex		Doctor ever told you had: arthritis		
		Total	1=Yes	2=No
Total	SAMSIZE	5249	2441	2808
	POPSIZE	29996716	13641419	16355297
	Row Percent	100.00	45.48	54.52
	SE Row Percent	0.000	0.905	0.905
1=Male	SAMSIZE	2443	934	1509
	POPSIZE	12631184	4660352	7970831
	Row Percent	100.00	36.90	63.10
	SE Row Percent	0.000	1.341	1.341
2=Female	SAMSIZE	2806	1507	1299
	POPSIZE	17365532	8981066	8384466
	Row Percent	100.00	51.72	48.28
	SE Row Percent	0.000	1.283	1.283

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

Exhibit 4. Stratum-Specific Hypothesis Tests for HSSEX*HAC1A

Variance Estimation Method: Taylor Series (WR)
Association Between GENDER and ARTHRITIS, Controlling for AGE (ADULTS 17+)

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

for: AGEGRP4 = Total.

Hypothesis Test Test Statistic	DF	Test Value	P-Value
CHISQ (Obs - Exp) Wald-F	1	131.43	0.0000

Exhibit 4. Stratum-Specific Hypothesis Tests for HSSEX*HAC1A-cont.

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

for: AGEGRP4 = 17-34.

Hypothesis Test Test Statistic	DF	Test Value	P-Value
CHISQ (Obs - Exp) Wald-F	1	2.31	0.1352

Exhibit 4. Stratum-Specific Hypothesis Tests for HSSEX*HAC1A-cont.

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

for: AGEGRP4 = 35-49.

Hypothesis Test Test Statistic	DF	Test Value	P-Value
CHISQ (Obs - Exp) Wald-F	1	7.33	0.0093

Exhibit 4. Stratum-Specific Hypothesis Tests for HSSEX*HAC1A-cont.

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

for: AGEGRP4 = 50-64.

Hypothesis Test Test Statistic	DF	Test Value	P-Value
CHISQ (Obs - Exp) Wald-F	1	45.39	0.0000

Exhibit 4. Stratum-Specific Hypothesis Tests for HSSEX*HAC1A-cont.

```
Test Statistics for Stratum-Specific Hypotheses
Variable HSSEX by Variable HAC1A

for: AGEGRP4 = 65-90+.
---------------------------------------------------------------
Hypothesis Test
   Test Statistic                  DF   Test Value      P-Value
---------------------------------------------------------------
CHISQ (Obs - Exp)
   Wald-F                           1        56.95       0.0000
---------------------------------------------------------------
```

Combining all four age groups, the Wald-*F* value for the CHISQ hypothesis of 131.43 (see ***Exhibit 4***) for testing an association between gender and arthritis is the same value obtained in ***Example 1*** with no adjustment for age. The CHISQ hypothesis at each level of age group shows that males and females differ significantly on the prevalence of arthritis except for the youngest age group, and the tables above show that females have the higher prevalence.

Exhibit 5. Stratum-Adjusted Hypothesis Test for HSSEX*HAC1A, Controlling for AGEGRP4

```
Variance Estimation Method: Taylor Series (WR)

Association Between GENDER and ARTHRITIS, Controlling for AGE (ADULTS 17+)

  Test Statistics for Stratum-Adjusted Hypotheses
  Variable HSSEX by Variable HAC1A
  Controlling for: Variable AGEGRP4

---------------------------------------------------------------
Hypothesis Test
   Test Statistic                  DF   Test Value      P-Value
---------------------------------------------------------------
CMH General Association
   Wald-F                           1        82.98       0.0000
---------------------------------------------------------------
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, ADULTS (17+)
```

The null hypothesis of the Cochran-Mantel-Haenszel (CMH) test for general association is that gender and arthritis are statistically independent, after controlling for age (***Exhibit 5***). The "Observed" and "Expected" calculations for the two-way table of gender with arthritis are estimated at each age level and then summed over the four age levels. The null hypothesis is rejected, indicating that males and females differ significantly on arthritis prevalence, after controlling for age. Note that the SUDAAN heading says "Controlling for: AGEGRP4." This means that the analysis is stratified by age (*i.e.*, age is controlled in the analysis). AGEGRP4 is an analytic stratification variable in this analysis, not a stratification variable in the sampling plan for NHANES III.