CROSSTAB Example #7

SUDAAN Statements and Results Illustrated

- Goodness-of-fit (GOF) Test
- GOF Test using GOFIT statement
- Wald-F (WALDF) Test
- Satterthwaite-adjusted Chi-square (SATADJCHI) Test
- SUBPOPX statement

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Using the NHANES III data, gives an example of the goodness-of-fit hypothesis test using the WALD-F and the Satterthwaite-adjusted chi-square test statistics.

Solution

This example illustrates the goodness-of-fit (GOF) hypothesis test in CROSSTAB. The primary data set consists of adults aged 18 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 in this example is displayed in **Exhibit 1**.

This example also uses external information from a more recent NHIS study to conduct a goodness-of-fit test using the GOFIT statement. According to the 2003-2005 NHIS survey, 21.6% of the adult population (defined as anyone $\geq$ 18 years of age) have been diagnosed with arthritis (see the website: *http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5540a2.htm*).

In the NHANES III data, there are 19,618 sample members age 18 and over (the SUBPOPX statement subsets the file to ages 18 and over). The GOFIT statement requests that the external NHIS percentages (21.6%=*Yes* vs. 78.4%=*No*) be compared to those in NHANES-III to see if the two prevalence distributions are statistically equivalent. To test the GOF hypothesis, both the default Wald-F (WALDF) and the Satterthwaite-adjusted chi-square (SATADJCHI) test statistics are requested.

The TABLES statement requests a 1-way table for HAC1A. This statement is not required in order to test the GOF hypothesis, but it does provide estimates from the current dataset. The GOFIT statement only requires that the variable(s) appear on the CLASS statement.

The PRINT statement is used in this example to request only the statistics of interest relating to estimates from the 1-way table and the GOF test.

The RFORMAT statements associate a SAS format with the CLASS variable.

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;
nest sdpstra6 sdppsu6;
weight wtpfqx6;

subpopx hsageir >=18;  /* ages 18 and over */
class hac1a;
table hac1a;  /* 1-way table of arthritis */
gofit hac1a=(21.6 78.4) / waldf satadjchi;

setenv labwidth=25;
print nsum rowper serow / gof=all;
RFORMAT HAC1A yn.;
RTITLE "GOFIT: Compare NHANES-III to NHIS (21.6 Percent of Adults Diagnosed with
        Arthritis)";
```

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  :  19618    Weighted count :184017835
Denominator degrees of freedom :     49
```

In the NHANES III data, there are 19,618 sample members aged 18 and over (*Exhibit 2*).

Below is the sample frequency distribution for the CLASS variable HAC1A (***Exhibit 3***).

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               4289       Yes
Ordered
  Position:
  2              15325        No
-------------------------------
```

Below are the three statistics NSUM, ROWPER, and SEROW (sample size, row percentage, and standard error) requested on the PRINT statement (***Exhibit 4***). This is where we learn that the NHANES III percentage of adults 18 and over diagnosed with arthritis is estimated to be 17.71%.

Exhibit 4. HAC1A Tabulation

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: HSAGEIR >= 18

GOFIT: Compare NHANES-III to NHIS (21.6 Percent of Adults Diagnosed with Arthritis)

by: Doctor ever told you had: arthritis.

-----------------------------------------------------------------------
|               |                 | Doctor ever told you had:     |
|               |                 | arthritis                     |
|               |                 |-------------------------------|
|               |                 | Total   | Yes     | No        |
-----------------------------------------------------------------------
|               |                 |         |         |           |
|               | Sample Size     |   19614 |    4289 |    15325  |
|               | Row Percent     |  100.00 |   17.71 |    82.29  |
|               | SE Row Percent  |    0.00 |    0.52 |     0.52  |
-----------------------------------------------------------------------
```

Below (***Exhibit 5***) is the GOF test for NHIS (21.6% "yes") vs. NHANES III (17.7% "yes") using the default Wald-*F* and the Satterthwaite-adjusted chi-square. Both tests are statistically significant (and equivalent, since there are only 2 categories in the distribution), which indicates that the two arthritis distributions (NHANES-III vs. the more recent NHIS) are statistically different.

Exhibit 5.   GOF Hypothesis Test

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: HSAGEIR >= 18

GOFIT: Compare NHANES-III to NHIS (21.6 Percent of Adults Diagnosed with Arthritis)

  Test Statistics for Goodness-of-Fit Hypotheses
  For: HAC1A = (0.216 0.784)

by: Test Statistic.
-----------------------------------------------------------------
Test Statistic                          Test
                        DF     Adj DF   Value      P-Value
-----------------------------------------------------------------
Wald-F                   1       .      55.7362    0.0000
Satterthwaite-adj chi-sq 1      1.00    55.7362    0.0000
-----------------------------------------------------------------
```