CROSSTAB Example #3

SUDAAN Statements and Results Illustrated

- Breslow-Day Test
- Odds Ratio or Relative Risk
- Breslow-Day Test of Homogeneity of "Odds Ratios"
- Prevalence Ratio
- Risk Statement

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Using NHANES III data, estimate the stratum-specific and stratum-adjusted odds ratio and prevalence ratio (or relative risk) for arthritis among adults (females vs. males).

Test the hypothesis of homogeneity of odds ratios via the Breslow-Day Test.

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 SUDAAN code for this example is displayed in ***Exhibit 1***.

When SUDAAN estimates the odds ratio or relative risk, it compares the 1st level of the row (or exposure) variable to the 2nd level of the row variable. Using HSSEX as the row variable would result in men being compared to women (with women the reference group). Because we wanted women compared to men (with men the reference group), we formed a new variable SEX_REC (1=female, 2=male) for this analysis.

The 2x2 table to be used for estimation of odds ratios and relative risks is defined by the last two variables on the TABLES statement—gender (SEX_REC) by arthritis (HAC1A). The odds ratio calculation in SUDAAN assumes that the column variable (disease) is coded as 1=disease and 2=no disease; this is how HAC1A is coded. The three-way cross-classification on the TABLES statement and the inclusion of the RISK statement requests SUDAAN to estimate odds ratios and relative risks for each level of the age group variable, as well as adjusted for age (common odds ratios and relative risks).

The RISK statement specifies the risk parameters of interest: stratum-specific odds ratios and relative risks for having arthritis (keywords *OR* and *RR1*), as well as the Mantel-Haenszel and Logit common odds ratios and common relative risks for having arthritis (keywords *MHOR, LOR, MHRR1* and *LRR1*).

The BDTEST statement instructs SUDAAN to use the stratified 2x2 table request on the TABLES statement and provide the Breslow-Day test for homogeneity of odds ratios across the strata (age groups). The ALL option tells SUDAAN to use all available test statistics for testing this hypothesis. Since DEFT1 is specified on the PROC statement, the Satterthwaite tests are included in the available test statistics.

The PRINT statement instructs SUDAAN to print point estimates and 95% confidence limits for the stratum-specific (STRRISK=default) and common (ADJRISK=default) odds ratios and relative risks specified on the RISK statement. Point estimates and confidence limits are the default statistics in the STRRISK and ADJRISK output groups. *RR1* (as opposed to *RR2*) indicates that the outcome variable of interest for arthritis is column 1 or code 1 (i.e., have arthritis). *RR1* will be the ratio of the female to male arthritis prevalence. Note that most epidemiologists would call this calculation a "prevalence ratio" rather than a relative risk.

The PRINT statement also requests the default statistics from the BTEST group (Breslow-Day statistics), which include hypothesis degrees of freedom, test statistic value, and *p*-value for each of the 5 test statistics.

The PRINT statement also requests the sample size (*NSUM*), estimated population size (*WSUM*), and row percentages (*ROWPER*) for each of the 2x2 (sex-by-arthritis) tables sub-classified by the levels of age group, as well as for all age groups combined (labeled "AGEGRP4=Total" in the first output table, on the following page).

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 yesno 1="1=Yes"
              2="2=No";
  value age 1="17-34"
            2="35-49"
            3="50-64"
            4="65-90+";
  value sexf 1="1=Female"
             2="2=Male";

PROC CROSSTAB  DATA=in.hanes3s3 FILETYPE=SAS DESIGN=WR deft1;
  NEST SDPSTRA6 SDPPSU6;
  WEIGHT WTPFQX6;

  SUBPOPX hsageir ge 17 / name=”Age 17+”;
  CLASS AGEGRP4 SEX_REC HAC1A;
  TABLES AGEGRP4*SEX_REC*HAC1A;
  RISK OR RR1 MHOR MHRR1 LOR LRR1;
  BDTEST / all;

  SETENV ROWWIDTH=12 COLWIDTH=10 DECWIDTH=2 LABWIDTH=28;
  PRINT NSUM="SAMSIZE" WSUM="POPSIZE" ROWPER / WSUMFMT=F9.0 NSUMFMT=F9.0;

  SETENV COLWIDTH=9 DECWIDTH=3 LABWIDTH=25;
  PRINT / ADJRISK=DEFAULT STRRISK=DEFAULT;

  SETENV COLWIDTH=10 DECWIDTH=2 LABWIDTH=28;
  PRINT / BTEST=DEFAULT BPVALFMT=F7.4 BDFFMT=F6.0;

  rformat agegrp4 age.;
  rformat hac1a yesno.;
  rformat sex_rec sexf.;
  RTITLE "Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association
          Between Gender and 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        May 2011
                              Release 11.0.0


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  :  20050    Weighted count :187647206
Denominator degrees of freedom :     49
```

The above SAS *.LST file in ***Exhibit 2*** shows that SUDAAN read in 20,050 adults in the data set, and all records met the SUBPOPX condition of adults age 17 and over. 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 (only the frequencies for CLASS variable *SEX_REC* is displayed here in ***Exhibit 3***, see ***Example 1*** for the other frequencies).

Exhibit 3. CLASS Variable Frequencies

```
Frequencies and Values for CLASS Variables
by: SEX_REC.
-------------------------------------
SEX_REC         Frequency       Value
-------------------------------------
Ordered
  Position:
  1                10649    1=Female
Ordered
  Position:
  2                 9401      2=Male
-------------------------------------
```

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

Exhibit 4. AGEGRP4*SEX_REC*HAC1A Crosstabulation

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Age 17+

Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association
Between Gender and Arthritis

by: AGEGRP4, SEX_REC, Doctor ever told you had: arthritis.

for: AGEGRP4 = Total.

-------------------------------------------------------------------------
|              |                  | Doctor ever told you had: arthritis  |
| SEX_REC      |                  |--------------------------------------|
|              |                  | Total      | 1=Yes      | 2=No       |
-------------------------------------------------------------------------
|              |                  |            |            |            |
| Total        | SAMSIZE          |      20046 |       4298 |      15748 |
|              | POPSIZE          |  187611487 |   32666641 |  154944847 |
|              | Row Percent      |     100.00 |      17.41 |      82.59 |
-------------------------------------------------------------------------
|              |                  |            |            |            |
| 1=Female     | SAMSIZE          |      10647 |       2728 |       7919 |
|              | POPSIZE          |   97980668 |   20877167 |   77103501 |
|              | Row Percent      |     100.00 |      21.31 |      78.69 |
-------------------------------------------------------------------------
|              |                  |            |            |            |
| 2=Male       | SAMSIZE          |       9399 |       1570 |       7829 |
|              | POPSIZE          |   89630819 |   11789474 |   77841345 |
|              | Row Percent      |     100.00 |      13.15 |      86.85 |
-------------------------------------------------------------------------
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, ADULTS (17+)
```

The above table (***Exhibit 4***) is the overall *SEX_REC* (gender) by *HAC1A* (arthritis) 2x2 table, for *AGEGRP4*=Total.

The following four 2 x 2 tables (gender by arthritis) are used to estimate the prevalence ratio and odds ratio of arthritis, females compared to males, for each level of age group.

Exhibit 4. AGEGRP4*SEX_REC*HAC1A Crosstabulation-cont.

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Age 17+

Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association Between Gender and Arthritis

by: AGEGRP4, SEX_REC, Doctor ever told you had: arthritis.

for: AGEGRP4 = 17-34.

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

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

For adults aged 17-34 years (***Exhibit 4***), the estimated prevalence ratio for arthritis (denoted relative risk or RR1 by SUDAAN) is (4.56) / (3.30) = 1.38. Hence, the arthritis prevalence for females is 38% higher than for males in this age group.

The estimated odds of arthritis for females aged 17-34 is (1638156) / (34296760) = .0478 and for males aged 17-34 is (1184692) / (34737872) = .0341. Therefore, the estimated odds ratio (denoted OR or odds ratio by SUDAAN) is (.0478) / (.0341) = 1.40. Females have 40% higher odds of arthritis than do males. Alternatively, the OR is estimated by [(1638156) x (34737872)] / [(34296760) x (1184692)].

Note that the prevalence ratio and odds ratio calculations are done on the estimated number of people in the population, and not on sample size figures. These results are displayed in the summary output at the end of this example.

Exhibit 4. AGEGRP4*SEX_REC*HAC1A Crosstabulation-cont.

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Age 17+

Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association Between Gender and Arthritis

by: AGEGRP4, SEX_REC, Doctor ever told you had: arthritis.

for: AGEGRP4 = 35-49.

SEX_REC		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
1=Female	SAMSIZE	2427	377	2050
	POPSIZE	27722459	3940978	23781481
	Row Percent	100.00	14.22	85.78
2=Male	SAMSIZE	2069	180	1889
	POPSIZE	25920111	2706268	23213843
	Row Percent	100.00	10.44	89.56

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

Applying the same calculations as above for adults aged 35-49 years, the estimated prevalence ratio is 1.36 (using results from ***Exhibit 4*** above) and the estimated odds ratio is 1.42, as indicated in the summary output at the end of this example.

Exhibit 4. AGEGRP4*SEX_REC*HAC1A Crosstabulation-cont.

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Age 17+

Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association Between Gender and Arthritis

by: AGEGRP4, SEX_REC, Doctor ever told you had: arthritis.

for: AGEGRP4 = 50-64.

SEX_REC		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
1=Female	SAMSIZE	1776	692	1084
	POPSIZE	16957761	6316967	10640794
	Row Percent	100.00	37.25	62.75
2=Male	SAMSIZE	1625	380	1245
	POPSIZE	15156961	3238161	11918800
	Row Percent	100.00	21.36	78.64

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

The estimated prevalence ratio is 1.74 and the estimated odds ratio is 2.19 for adults aged 50-64 years, as indicated in the summary output at the end of this example (***Exhibit 4***).

Exhibit 4. AGEGRP4*SEX_REC*HAC1A Crosstabulation-cont.

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Age 17+

Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association Between Gender and Arthritis

by: AGEGRP4, SEX_REC, Doctor ever told you had: arthritis.

for: AGEGRP4 = 65-90+.

SEX_REC		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
1=Female	SAMSIZE	2806	1507	1299
	POPSIZE	17365532	8981066	8384466
	Row Percent	100.00	51.72	48.28
2=Male	SAMSIZE	2443	934	1509
	POPSIZE	12631184	4660352	7970831
	Row Percent	100.00	36.90	63.10

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

The estimated prevalence ratio is 1.40 and the estimated odds ratio is 1.83 for adults aged 65-90 years, as indicated in the summary output at the end of this example (***Exhibit 4***).

Exhibit 5. Stratum-Specific Odds Ratios and Relative Risks

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Age 17+

Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association Between
  Gender and Arthritis

Stratum Specific Odds Ratios and Relative Risks
   Variable SEX_REC by Variable HAC1A

by: AGEGRP4, Stratum Specific Risk.
------------------------------------------------------------
AGEGRP4                                 Lower 95%   Upper 95%
   Stratum Specific Risk        Value   Limit       Limit
------------------------------------------------------------
17-34
   Odds Ratio                   1.401       0.882       2.223
   Relative Risk Col 1          1.382       0.886       2.158
35-49
   Odds Ratio                   1.421       1.087       1.859
   Relative Risk Col 1          1.362       1.074       1.725
50-64
   Odds Ratio                   2.185       1.722       2.772
   Relative Risk Col 1          1.744       1.467       2.073
65-90+
   Odds Ratio                   1.832       1.558       2.155
   Relative Risk Col 1          1.402       1.278       1.537
------------------------------------------------------------
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, ADULTS (17+)
```

The above output (***Exhibit 5***) confirms our previous calculations for the four age-specific odds ratios and relative risks. Note that the 95% confidence intervals on the prevalence ratios overlap, so the prevalence ratio may not vary substantially over the four age groups. All age-specific confidence intervals on the prevalence ratio exclude 1.0 except for the youngest age group; this finding is consistent with the age-specific CHISQ tests of the previous ***Example 2***. Similarly, the odds ratio may not vary substantially over the four age groups; this can be tested via logistic regression using either LOGISTIC or MULTILOG, or via the Breslow-Day test for homogeneity of odds ratios, shown next in ***Exhibit 6***.

Exhibit 6. Test of the Breslow-Day Hypothesis of Homogeneity of Odds Ratios

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Age 17+

Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association Between
  Gender and Arthritis

  Test Statistics for Breslow-Day Hypothesis
  Variable SEX_REC by Variable HAC1A
  Controlling for: Variable AGEGRP4

by: Hypothesis Test, Test Statistic.
------------------------------------------------------------------------
Hypothesis Test
   Test Statistic                  DF       Adj DF   Test Value   P-Value
------------------------------------------------------------------------
Breslow-Day: Homogeneity of
  Odds Ratios
   Wald chi-square                 3          .           6.93    0.0742
   Wald-F                          3          .           2.31    0.0879
   Adj Wald F                      3          .           2.22    0.0987
   Satterthwaite-adj chi-sq        3       2.55           6.15    0.0747
   Satterthwaite-adj F             3       2.55           2.42    0.0869
------------------------------------------------------------------------
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, ADULTS (17+)
```

The above output (***Exhibit 6***) provides the test of the Breslow-Day hypothesis of homogeneity of odds ratios across analysis strata (age group in our example). All five test statistics were requested via the ALL option on BDTEST. P-values ranged from 0.07 to 0.10. Since the odds ratios across age groups do not vary significantly, we proceed with estimating the stratum-adjusted odds ratios in ***Exhibit 7***.

Exhibit 7. Stratum-Adjusted Odds Ratios and Relative Risks

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Age 17+

Prevalence Ratio, Odds Ratio, and Breslow-Day Test Measuring Association Between
  Gender and Arthritis

Summary Statistics for Variable SEX_REC by Variable HAC1A
Controlling for: Variable AGEGRP4

by: Adjusted Risk.
------------------------------------------------------------
Adjusted Risk                          Lower 95%   Upper 95%
                              Value    Limit       Limit
------------------------------------------------------------
MH Common OR                  1.746       1.526       1.997
MH Common RR Col 1            1.483       1.348       1.631
Logit Common OR               1.724       1.525       1.949
Logit Common RR Col 1         1.418       1.309       1.536
------------------------------------------------------------
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE, ADULTS (17+)
```

Exhibit 7 presents the common (age-adjusted) odds ratios and relative risks for having arthritis, via both Mantel-Haenszel and Logit approaches. The Mantel-Haenszel and Logit common (age-adjusted) odds ratios are 1.75 and 1.72, respectively. The Mantel-Haenszel and Logit common relative risks for column 1 (having arthritis) are 1.48 and 1.42, respectively. These results quantify the increase in arthritis risk for females vs. males, adjusted for age (note the *Controlling for: Variable AGEGRP4* line in the output above).

The Mantel-Haenszel vs. Logit 95% confidence intervals for the age-adjusted odds ratios (and also for the age-adjusted relative risks) overlap, showing consistency across methods. Also, all confidence intervals on the age-adjusted risk estimates exclude the value 1.0; this is consistent with the Cochran-Mantel-Haenszel age-adjusted test for general association between sex and arthritis (p=0.0000, ***Example 2***). Further covariate analysis on the common odds ratio could also be investigated via logistic regression using either the MULTILOG or LOGISTIC procedures.