RATIO Example #1

SUDAAN Statements and Results Illustrated

- NEWVAR
- DENOM
- NUMER
- SETENV
- WEIGHT

Input Data Set(s): NHANES3S3.SAS7bdat

Example

Using NHANES III, estimate the proportion of DMF (decayed, missing, filled) teeth among adults that result from disease. Estimate for the entire population and for age groups by decade.

This example also demonstrates the use of the NEWVAR statement in variable creation.

Solution

The data set is adults aged 17+ from NHANES III, a cross-sectional sample survey of the civilian, non-institutionalized population aged two months or older and fielded during 1988-1994. The dental variables in this example are from the MEC (mobile examination center) physical examination, which was conducted subsequent to the home interview component of NHANES III. The appropriate sample weight variable is WTPFEX6, since MEC examined variables are analyzed and because all six years of data are analyzed. The appropriate stratification and PSU variables for analysis of six years of data are SDPSTRA6 and SDPPSU6, respectively.

The SUBPOPX statement (***Exhibit 1*** and ***Exhibit 5***) restricts the analysis to MEC examined adults who had some part of the dental exam (17,235 out of 17,705). The dental exam flag variable (DEPEXFLG) is used on the SUBPOPX statement; a code of 1 indicates some part of the dental exam was done.

This analysis uses two DMF teeth measurements: total DMF teeth (DEPDMFT2) and DMF teeth due to disease (DEPDMFT1). PROC RATIO (***Exhibit 5***) estimates the proportion of total DMF teeth due to disease by specifying total DMF teeth (DEPDMFT2) as the denominator variable (DENOM statement) and DMF teeth due to disease (DEPDMFT1) as the numerator variable (NUMER statement).

It is desired to estimate this ratio parameter (a proportion) for different age groups as well as for the total adult population. The NEWVAR statement forms a recode of the variable HSAGEIR (age in years) into seven age groups: 17-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80+. NEWVAR is the first statement implemented by SUDAAN, regardless of its order in the procedure code. Thus, AGEGRP appears on the CLASS statement, and SUDAAN treats this as a categorical variable with seven levels: 1, 2, 3, 4, 5, 6, 7 (***Exhibit 5***). To ensure that the variable AGEGRP was created as we had planned, we include a CROSSTAB to display the cross-tabulation of HSAGEIR with AGEGRP (***Exhibit 1***). Both HSAGEIR and AGEGRP are re-labelled in the output using the RLABEL statement.

This example was run in SAS-Callable SUDAAN, and the programming code is presented in ***Exhibit 1*** and ***Exhibit 5***. Note that the basic SUDAAN code is the same for both Standalone and SAS-Callable versions.

Exhibit 1. SAS-Callable SUDAAN Code: CROSSTAB

```
libname in "\\rtints29\sudaan\data\nhanes3";
options linesize=95 pagesize=60 nocenter;

proc format;
  value age 1="1=17-29"
            2="2=30-39"
            3="3=40-49"
            4="4=50-59"
            5="5=60-69"
            6="6=70-79"
            7="7=80+";

PROC CROSSTAB DATA=in.HANES3S3 FILETYPE=SAS DESIGN=WR;
  NEST   SDPSTRA6 SDPPSU6;
  WEIGHT WTPFEX6;

  SUBPOPX DEPEXFLG=1 / NAME="Had Dental Exam in MEC";

  NEWVAR AGEGRP:      If 17 le HSAGEIR le 29 then AGEGRP=1
                 elseif 30 le HSAGEIR le 39 then AGEGRP=2
                 elseif 40 le HSAGEIR le 49 then AGEGRP=3
                 elseif 50 le HSAGEIR le 59 then AGEGRP=4
                 elseif 60 le HSAGEIR le 69 then AGEGRP=5
                 elseif 70 le HSAGEIR le 79 then AGEGRP=6
                 else if HSAGEIR ge 80 then AGEGRP=7;
  CLASS HSAGEIR AGEGRP / include=missing;
  TABLES HSAGEIR*AGEGRP;

  setenv labwidth=10 colwidth=5 decwidth=0;
  PRINT NSUM / style=nchs;
  RLABEL HSAGEIR="Age in Years";
  RLABEL AGEGRP="Age Group";
  RFORMAT agegrp age.;
  RTITLE "Mapping HSAGEIR into AGEGRP";
```

Exhibit 1 contains the CROSSTAB code for checking the correct mapping of the continuous HSAGEIR (in years) into the 7-level AGEGRP variable. It uses the same survey design, SUBPOPX, and NEWVAR statements that will be included in RATIO. The purpose of the CROSSTAB run is mainly a quality assurance check of the creation of AGEGRP from HSAGEIR. For this purpose, we specified INCLUDE=MISSING on the CLASS statement in order to determine how potential missing values of HSAGEIR were assigned and whether the created variable has any missing values. We will not need to specify this option in RATIO.

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 11.0.0

DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a
With Replacement (WR) Design
    Sample Weight: WTPFEX6
    Stratification Variables(s): SDPSTRA6
    Primary Sampling Unit: SDPPSU6

Number of observations read    :  17705    Weighted count :187521389
Number of observations skipped :   2345
(WEIGHT variable nonpositive)
Observations in subpopulation  :  17235    Weighted count :182292666
Denominator degrees of freedom :     49
```

Exhibit 2 indicates that SUDAAN read in 20,050 observations from the data set (17,705 + 2,345). SUDAAN excluded from the analysis 2,345 adults who had a value of 0 for the weight variable WTPFEX6; these subjects participated in the home interview component of NHANES III but not in the physical examination component conducted at the MEC. Of the 17,705 MEC examined adults, 17,235 participated in at least part of the dental exam; this is the subpopulation that was analyzed.

Exhibit 2 also indicates that the denominator degrees of freedom (ddf) for NHANES III (six years of data) is 49 (i.e., 98 "pseudo-PSUs" minus 49 "pseudo-strata"). See ***Example 1*** for more discussion of NHANES III ddf.

The CROSSTAB results are presented next. The purpose of running CROSSTAB is to check the mapping of HSAGEIR into AGEGRP. The frequencies for the CLASS variable AGEGRP are displayed next (***Exhibit 3***). Frequencies for CLASS variable HSAGEIR were also produced, but are lengthy and not displayed here.

Exhibit 3. CROSSTAB: Frequencies for CLASS Variable AGEGRP

```
Frequencies and Values for CLASS Variables
by: Age Group.
------------------------------------
Age Group       Frequency      Value
------------------------------------
Ordered
  Position:
  1                  4541    1=17-29
Ordered
  Position:
  2                  3244    2=30-39
Ordered
  Position:
  3                  2513    3=40-49
Ordered
  Position:
  4                  1801    4=50-59
Ordered
  Position:
  5                  2236    5=60-69
Ordered
  Position:
  6                  1701    6=70-79
Ordered
  Position:
  7                  1199      7=80+
------------------------------------
```

There are 7 levels for AGEGRP (re-labeled Age Group), as seen above.

Exhibit 4. CROSSTAB Results (Checking the AGEGRP Recode)

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Had Dental Exam in MEC

Mapping HSAGEIR into AGEGRP

Sample Size

by: Age in Years, Age Group.

Age in Years	Age Group Total	1=17-29	2=30-39	3=40-49	4=50-59	5=60-69	6=70-79	7=80+
Total	17235	4541	3244	2513	1801	2236	1701	1199
17	402	402	0	0	0	0	0	0
18	357	357	0	0	0	0	0	0
19	348	348	0	0	0	0	0	0
20	334	334	0	0	0	0	0	0
21	334	334	0	0	0	0	0	0
22	343	343	0	0	0	0	0	0
23	378	378	0	0	0	0	0	0
24	376	376	0	0	0	0	0	0
25	327	327	0	0	0	0	0	0
26	331	331	0	0	0	0	0	0
27	348	348	0	0	0	0	0	0
28	335	335	0	0	0	0	0	0
29	328	328	0	0	0	0	0	0
30	347	0	347	0	0	0	0	0
31	356	0	356	0	0	0	0	0
32	331	0	331	0	0	0	0	0
33	333	0	333	0	0	0	0	0
34	325	0	325	0	0	0	0	0
35	320	0	320	0	0	0	0	0
36	335	0	335	0	0	0	0	0
37	309	0	309	0	0	0	0	0
38	325	0	325	0	0	0	0	0
39	263	0	263	0	0	0	0	0
40	341	0	0	341	0	0	0	0
41	326	0	0	326	0	0	0	0
42	320	0	0	320	0	0	0	0
43	266	0	0	266	0	0	0	0
44	250	0	0	250	0	0	0	0
45	229	0	0	229	0	0	0	0
46	216	0	0	216	0	0	0	0
47	208	0	0	208	0	0	0	0
48	183	0	0	183	0	0	0	0
49	174	0	0	174	0	0	0	0
50	195	0	0	0	195	0	0	0
51	165	0	0	0	165	0	0	0
52	188	0	0	0	188	0	0	0
53	177	0	0	0	177	0	0	0
54	188	0	0	0	188	0	0	0

Exhibit 4. CROSSTAB Results (AGEGRP Recode - cont'd.)

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Had Dental Exam in MEC

Mapping HSAGEIR into AGEGRP

Sample Size

by: Age in Years, Age Group.
-------------------------------------------------------------------------
Age in        Age Group
  Years             1=17-   2=30-   3=40-   4=50-   5=60-   6=70-
              Total  29      39      49      59      69      79      7=80+
-------------------------------------------------------------------------
55              189       0       0       0     189       0       0       0
56              194       0       0       0     194       0       0       0
57              180       0       0       0     180       0       0       0
58              187       0       0       0     187       0       0       0
59              138       0       0       0     138       0       0       0
60              239       0       0       0       0     239       0       0
61              251       0       0       0       0     251       0       0
62              221       0       0       0       0     221       0       0
63              230       0       0       0       0     230       0       0
64              240       0       0       0       0     240       0       0
65              232       0       0       0       0     232       0       0
66              206       0       0       0       0     206       0       0
67              215       0       0       0       0     215       0       0
68              199       0       0       0       0     199       0       0
69              203       0       0       0       0     203       0       0
70              214       0       0       0       0       0     214       0
71              212       0       0       0       0       0     212       0
72              215       0       0       0       0       0     215       0
73              210       0       0       0       0       0     210       0
74              178       0       0       0       0       0     178       0
75              162       0       0       0       0       0     162       0
76              157       0       0       0       0       0     157       0
77              143       0       0       0       0       0     143       0
78              109       0       0       0       0       0     109       0
79              101       0       0       0       0       0     101       0
80              207       0       0       0       0       0       0     207
81              198       0       0       0       0       0       0     198
82              147       0       0       0       0       0       0     147
83              126       0       0       0       0       0       0     126
84              111       0       0       0       0       0       0     111
85               91       0       0       0       0       0       0      91
86               84       0       0       0       0       0       0      84
87               62       0       0       0       0       0       0      62
88               42       0       0       0       0       0       0      42
89               33       0       0       0       0       0       0      33
90               98       0       0       0       0       0       0      98
-------------------------------------------------------------------------
```

Exhibit 4 indicates that all values of HSAGEIR in the subpopulation are mapped into the intended cells of the 7-level AGEGRP. There were no missing values for HSAGEIR or AGEGRP in the sub-population.

Exhibit 5. RATIO Code

```
PROC RATIO DATA=in.HANES3S3 FILETYPE=SAS DESIGN=WR;
  NEST   SDPSTRA6 SDPPSU6;
  WEIGHT WTPFEX6;

  SUBPOPX DEPEXFLG=1 / NAME="Had Dental Exam in MEC";

  NUMER  DEPDMFT1;     /* # DMF TEETH DUE TO DISEASE */
  DENOM  DEPDMFT2;     /* # DMF TEETH DUE TO ANY CAUSE */

  NEWVAR AGEGRP:     If 17 le HSAGEIR le 29 then AGEGRP=1
                 elseif 30 le HSAGEIR le 39 then AGEGRP=2
                 elseif 40 le HSAGEIR le 49 then AGEGRP=3
                 elseif 50 le HSAGEIR le 59 then AGEGRP=4
                 elseif 60 le HSAGEIR le 69 then AGEGRP=5
                 elseif 70 le HSAGEIR le 79 then AGEGRP=6
                 else if HSAGEIR ge 80 then AGEGRP=7;
  CLASS AGEGRP;
  TABLES AGEGRP;

  SETENV LABWIDTH=20 COLSPCE=1 colwidth=10 decwidth=0;
  PRINT NSUM="Sample" WSUM="PopSize" WYSUM="Disease DMFT" WXSUM="Any Cause DMFT"
        RHAT="Ratio" / NSUMFMT=F6.0 WSUMFMT=F9.0 RHATFMT=F6.4 STYLE=NCHS;

  SETENV LABWIDTH=20 COLSPCE=1 colwidth=6 decwidth=4;
  PRINT RHAT="Ratio" SERHAT="SE" LOWRHAT="Lower 95% Limit" UPRHAT="Upper 95% Limit"
        / STYLE=NCHS;

  RLABEL agegrp="Age at Interview GROUPED";
  RLABEL depdmft1="Disease";
  RLABEL depdmft2="Any Cause";
  RFORMAT agegrp age.;
  RTITLE "PROPORTION OF DMF TEETH DUE TO DISEASE, BY 7 AGE GROUPS"
         "U.S. ADULTS AGED 17+ YEARS";
  RFOOTNOTE "NHANES-III, 1988-1994, JULY 1997 DATA RELEASE";
```

The TABLES statement requests the ratio parameter to be estimated for each of the seven age groups. Two PRINT statements were used because there were too many statistics requested to fit on a single page. Each PRINT statement produces a table of results.

Exhibit 6. RATIO Results

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Had Dental Exam in MEC

PROPORTION OF DMF TEETH DUE TO DISEASE, BY 7 AGE GROUPS
U.S. ADULTS AGED 17+ YEARS

Variable Age at Interview GROUPED	Sample	PopSize	Disease DMFT	Any Cause DMFT	Ratio
Disease/Any Cause					
Total	17219	182175161	2490153345	2552550674	0.9756
1=17-29	4535	48463261	318526762	340455640	0.9356
2=30-39	3241	41094542	460963034	478129690	0.9641
3=40-49	2508	32337038	482029945	491416760	0.9809
4=50-59	1800	21365215	395329851	401780915	0.9839
5=60-69	2236	19568863	399886504	405105060	0.9871
6=70-79	1700	13464941	295717760	297207091	0.9950
7=80+	1199	5881302	137699488	138455519	0.9945

NHANES-III, 1988-1994, JULY 1997 DATA RELEASE

Exhibit 6. RATIO Results (cont'd.)

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: Had Dental Exam in MEC

PROPORTION OF DMF TEETH DUE TO DISEASE, BY 7 AGE GROUPS
U.S. ADULTS AGED 17+ YEARS

Variable Age at Interview GROUPED	Ratio	SE	Lower 95% Limit	Upper 95% Limit
Disease/Any Cause				
Total	0.9756	0.0015	0.9726	0.9785
1=17-29	0.9356	0.0044	0.9267	0.9445
2=30-39	0.9641	0.0030	0.9580	0.9702
3=40-49	0.9809	0.0017	0.9774	0.9844
4=50-59	0.9839	0.0029	0.9782	0.9897
5=60-69	0.9871	0.0034	0.9803	0.9939
6=70-79	0.9950	0.0011	0.9929	0.9971
7=80+	0.9945	0.0018	0.9910	0.9981

NHANES-III, 1988-1994, JULY 1997 DATA RELEASE

Of the 17,235 adults who participated in some part of the dental exam, almost all (17,219) had values for the two dental variables in this analysis (see Total row in ***Exhibit 6***). These 17,219 sample adults represent 182,175,161 adults in the population. The total number of DMF teeth among these 182 million adults is estimated to be 2,552,550,674 (2.55 billion). The total number of DMF teeth due to disease among the 182 million adults is estimated to be 2.49 billion. The proportion of DMF teeth due to disease is estimated to be 2,490,153,345/2,552,550,674 (0.9756). The estimated standard error for the point estimate 0.9756 is 0.0015, with a 95% confidence interval of (0.9726, 0.9785).

The seven rows below the Total row in ***Exhibit 6*** are the seven age groups (roughly by decade). The same quantities as given for the Total row are given for each age decade. The estimated proportion of total DMF teeth due to disease increases with age, with a slight decrease in the last decade (age 80+).