

# DESCRIPT Example #8

## *SUDAAN Statements and Results Illustrated*

- TOTPER option
- NSUM option
- WEIGHT
- LEVELS
- SUBGROUP

## *Input Data Set(s): NHANES3S3.SAS7bdat*

### *Example*

*Compare male and female adults on arthritis prevalence, within age group, using NHANES III.*

### *Solution*

*Example 6* showed that race/ethnicity seems to be related to arthritis, with a lower arthritis prevalence for Mexican-Americans and “other” (both around 10%), compared to 19% for non-Hispanic whites and 16% for non-Hispanic blacks. Previous examples also showed that age and gender are strongly related to arthritis prevalence. Since the non-Hispanic white subpopulation is older, it likely is misleading to directly compare race/ethnicity groups on arthritis prevalence. The percentage of adults aged 17-34 years is 35% for non-Hispanic white, 45% for non-Hispanic black, 55% for Mexican-American, and 46% for “other.” The percentage of adults aged 65+ years is 18% for non-Hispanic white, 12% for non-Hispanic black, 6% for Mexican-American and 8% for “other.” Thus, we wish to estimate the prevalence of arthritis for each race/ethnicity group, but adjusted for age group and gender.

We need to choose a standard distribution for the cross-classification of age group and gender. We use the distribution given by adults in NHANES III, since this sample is post-stratified to the U.S. population. The CROSSTAB program below generates the standardized weights (*Exhibit 1*). The TABLES statement requests the age by gender percentage distribution for the entire adult population. Since we wish to know the percentage of the total population comprised by each age/gender combination, we request TOTPER on the PRINT statement.

This example was run in SAS-Callable SUDAAN, and the programming code is presented below. 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 sex 1="1=male"
           2="2=female";
  value age 1="1=17-34"
           2="2=35-49"
           3="3=50-64"
           4="4=65-90+";
  value race 1="1=nH_white"
            2="2=nH_black"
            3="3=Mex_Amer"
            4="4=other";
  value yesno 1="1=Yes";

PROC CROSSTAB DATA=in.HANES3S3 FILETYPE=SAS DESIGN=WR;
  NEST SDPSTRA6 SDPPSU6;
  WEIGHT WTPFQX6;

  SUBGROUP AGEGRP4 HSSEX;
  LEVELS 4 2;
  TABLES AGEGRP4*HSSEX;

  PRINT NSUM TOTPER / STYLE=NCHS NSUMFMT=F11.0 TOTPERFMT=F12.3;
  rformat hssex sex.;
  rformat agegrp4 age.;
  RTITLE "AGE/SEX DISTRIBUTION, ADULTS (17+), U.S.";
  RFOOTNOTE "NHANES-III, 1988-1994, JULY 1997 DATA RELEASE";
```

## 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: WTPFQX6
  Stratification Variables(s): SDPSTRA6
  Primary Sampling Unit: SDPPSU6

Number of observations read      : 20050      Weighted count :187647206
Denominator degrees of freedom : 49
```

**Exhibit 3. CROSSTAB Results**

```

Variance Estimation Method: Taylor Series (WR)
AGE/SEX DISTRIBUTION, ADULTS (17+), U.S.
by: AGEGRP4, Sex.
-----
AGEGRP4
Sex                Sample Size      Tot Percent
-----
Total
  Total                20050            100.000
  1=male                9401             47.769
  2=female              10649            52.231
1=17-34
  Total                6900             38.294
  1=male                3262             19.144
  2=female              3638             19.150
2=35-49
  Total                4496             28.587
  1=male                2069             13.813
  2=female              2427             14.774
3=50-64
  Total                3402             17.126
  1=male                1625             8.077
  2=female              1777             9.049
4=65-90+
  Total                5252             15.993
  1=male                2445             6.735
  2=female              2807             9.258
-----
NHANES-III, 1988-1994, JULY 1997 DATA RELEASE

```

The standardized weights for the age-by-gender distribution are given in *Exhibit 3*. Since the variables will be listed on the STDVAR statement with age first, followed by gender, the weights will be listed on the STDWGT statement in the order of sex, nested within age group, as in

(.1914 .1915 .1381 .1477 .0808 .0905 .0674 .0926).

*Exhibit 4* contains the SUDAAN program to generate the age/gender-adjusted arthritis prevalence for each race/ethnicity group. The VAR and CATLEVEL statements request that the percentage with arthritis be estimated. The TABLES statement requests the prevalence for each race/ethnicity group. The STDVAR statement requests that the prevalence be adjusted by the cross-classification of age group and gender. The STDWGT statement specifies the standard weights. Note that there are eight values on the STDWGT statement, since there are four levels of age group and two levels of gender.

## Exhibit 4. SAS-Callable SUDAAN Code: DESCRIPT

```

PROC DESCRIPT DATA=in.HANES3S3 FILETYPE=SAS DESIGN=WR;
  NEST SDPSTRA6 SDPPSU6;
  WEIGHT WTPFQX6;

  VAR HAC1A;
  CATLEVEL 1;

  SUBGROUP DMARETHN AGEGRP4 HSSEX;
  LEVELS 4 4 2;
  TABLES DMARETHN;
  STDVAR AGEGRP4 HSSEX;
  STDWGT .1914 .1915 .1381 .1477 .0808 .0905 .0673 .0926;

  SETENV COLWIDTH=7 decwidth=2 LABWIDTH=17 colspce=0;
  PRINT / NSUMFMT=f6.0 WSUMFMT=F9.0 TOTALFMT=F8.0 STYLE=NCHS;
  rformat dmarethn race.;
  rformat hac1a yesno.;
  RTITLE "AGE/SEX ADJUSTED PREVALENCE RATES FOR ARTHRITIS"
        "BY RACE/ETHNICITY, U.S. ADULTS AGED 17+ YEARS";
  RFOOTNOTE "NHANES-III, 1988-1994, JULY, 1997 DATA RELEASE";

```

## Exhibit 5. DESCRIPT Results: Standardization

Variance Estimation Method: Taylor Series (WR)  
Standardized estimates

AGE/SEX ADJUSTED PREVALENCE RATES FOR ARTHRITIS  
BY RACE/ETHNICITY, U.S. ADULTS AGED 17+ YEARS

by: Variable, Race-ethnicity.

Variable					SE	Lower	Upper
Race-					Percent	95%	95%
Ethnicity	Sample	Weighted	Total	Percent	Percent	Limit	Limit
	Size	Size				Percent	Percent
-----							
Doctor ever							
told you							
had							
arthritis:							
1=Yes							
Total	20046	187611487	32666641	17.42	0.34	16.74	18.11
1=nH_white	8480	142595429	26880246	17.78	0.48	16.84	18.76
2=nH_black	5485	20995070	3455547	19.19	0.58	18.05	20.39
3=Mex_Amer	5306	9827951	964747	15.65	0.70	14.28	17.11
4=other	775	14193038	1366101	12.15	1.47	9.50	15.42

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

The age/gender-adjusted prevalences are given in the Percent column of *Exhibit 5*, with their estimated standard error in the SE Percent column. The unadjusted (see *Example 6*) and age/sex adjusted (*Exhibit 5*) arthritis prevalence, respectively, for each race/ethnicity group are:

- Non-Hispanic white: 18.85% and 17.78%
- Non-Hispanic black: 16.46% and 19.19%
- Mexican-American: 9.82% and 15.64%
- "Other": 9.63% and 12.15%.

The effect of the adjustment was to lower the prevalence by 1% for non-Hispanic whites and to substantially increase (from 3% to 6%) the prevalence for all other race/ethnicity groups. The differences between the four race/ethnicity groups on unadjusted arthritis prevalence have been reduced by age/gender standardization. Most of this effect is due to age standardization.

In another DESCRIPT program, you could use the CONTRAST, DIFFVAR, or PAIRWISE statements to test the null hypothesis that the age/gender-adjusted arthritis prevalence is the same for the race-ethnicity groups. The size of the standard errors above suggests that some of the race/ethnicity groups would differ significantly on age/gender-adjusted arthritis prevalence.

The weighted size and total figures in the printout table above are the same figures as in *Example 6*. The ratio of these two figures yields the unadjusted prevalence for each race/ethnicity group.