

# RATIO Example #3

## *SUDAAN Statements and Results Illustrated*

- DENOM
- DENCAT
- NUMER
- NUMCAT
- SUBPOPX

## *Input Data Set(s): SADLTST3.SSD*

### *Example*

*Among adult female cancer survivors in the U.S., estimate the proportion who have had breast cancer, using the NHIS (National Health Interview Survey) data for 1997.*

### *Solution*

The NHIS is an annual, multi-purpose health sample survey conducted by the NCHS (National Center for Health Statistics).

In this example, we use the sample adult data file with 36,116 observations. The stratification and primary sampling unit variables are named STRATUM and PSU, respectively, and appear on the NEST statement below. The weight variable for the sample adult file is WTFA\_SA and appears on the WEIGHT statement. The PROC statement contains DESIGN=WR (i.e., unequal probability sampling of PSUs with replacement).

Each sample adult was asked if he/she had ever been told by a doctor or other health professional that he/she had cancer or a malignancy of any kind (the variable CANEV, 1=yes, 2=no). If yes, then the subject was asked what kind(s) of cancer. There are several variables for type of cancer, each coded as (1=yes, 2=no). The variable CNKIND5 refers to breast cancer. We use PROC RATIO to estimate the following proportion: (number of adult female cancer survivors, or patients, who have had breast cancer) / (number of adult female cancer survivors, or patients).

The denominator is estimated by the sum of the weight variable for sample female adults who have had cancer (i.e., the variable CANEV on the DENOM statement and the value 1 [for yes] on the DENCAT statement). The numerator is estimated by the sum of the weight variable for sample female adults who have had breast cancer (i.e., the variable CNKIND5 on the NUMER statement and the value 1 [for yes] on the NUMCAT statement).

The SUBPOPX statement restricts the analysis to females only (SEX = 2). The second part of the SUBPOPX statement excludes ten females who said they had cancer (CANEV = 1), but did not know or refused to say whether they ever had breast cancer (CNKIND5 = 7 for *don't know*, or 9 for *refuse*).

The PRINT statement changes the default label for most of the estimated quantities. The SETENV statement and the FMT options on the PRINT statement manipulate the appearance of the printout.

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

## Exhibit 1. SAS-Callable SUDAAN Code

```
libname in v604 "\\rtints29\sudaan\data\nhis1997";
options linesize=95 pagesize=60 nocenter;

PROC RATIO DATA=in.SADLTST3 DESIGN=WR FILETYPE=SAS NOMARG;
  NEST STRATUM PSU;
  WEIGHT WTFA_SA;

  SUBPOPX SEX=2 and CNKIND5 in(1,2) /
    NAME="FEMALES & ANSWER BRCA TYPE QX YES OR NO";

  NUMER CNKIND5; /* BREAST CANCER EVER? 1=YES */
  NUMCAT 1;
  DENOM CANEV; /* CANCER EVER? 1=YES */
  DENCAT 1;

  SETENV ROWWIDTH=22 decwidth=4 colwidth=10;
  PRINT NSUM="SAMSIZE" WSUM="POPSIZE" WXSUM="POP CANCER" WYSUM="POP BRCA"
    RHAT="PROP. BRCA" SERHAT="SE" LOWRHAT="Lower 95% Limit"
    UPRHAT="Upper 95% Limit" / NSUMFMT=F9.0 WSUMFMT=F9.0 WXSUMFMT=F10.0
    WYSUMFMT=F8.0;
  RTITLE "PERCENT OF ADULT FEMALE CANCER PATIENTS WHO HAD BREAST CANCER"
    "U.S., 1997";
  RFOOTNOTE "NHIS, 1997, SAMPLE ADULT DATASET";
```

## 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: WTFA_SA
  Stratification Variables(s): STRATUM
  Primary Sampling Unit: PSU

Number of observations read      : 36116      Weighted count :195276321
Observations in subpopulation   : 1521      Weighted count : 7202079
Denominator degrees of freedom  : 339
```

From *Exhibit 2*, SUDAAN read in 36,116 observations from the sample adult data set, and all observations had a positive value for the weight variable WTFA\_SA. The estimated adult population size in 1997 was 195 million. The 1,521 observations in the subpopulation make inference to an estimated 7.2 million female adult cancer survivors (or patients) in the U.S. in 1997; this estimate includes non-melanoma skin cancer. The denominator degrees of freedom (ddf) for the 1997 NHIS is 339 (*i.e.*, 678 “pseudo-PSUs” minus 339 “pseudo-strata”).

### Exhibit 3. RATIO Results

```

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: FEMALES & ANSWER BRCA TYPE QX YES OR NO

PERCENT OF ADULT FEMALE CANCER PATIENTS WHO HAD BREAST CANCER
U.S., 1997

-----
| Variable                |          | SUDAAN  |
|                          |          | Reserved|
|                          |          | Variable|
|                          |          | One     |
|                          |          |-----|
|                          |          | 1       |
|-----|-----|
| What kind of            | SAMSIZE  | 1521    |
| cancer...breast:       | POPSIZE  | 7202079 |
| 1/Ever told by a      | POP CANCER | 7202079 |
| doctor you had cancer: | POP BRCA  | 1974567 |
| 1                      | PROP. BRCA | 0.2742  |
|                        | SE        | 0.0133  |
|                        | Lower 95% Limit | 0.2480  |
|                        | Upper 95% Limit | 0.3004  |
|-----|-----|
NHIS, 1997, SAMPLE ADULT DATASET

```

**Exhibit 3** shows the estimated number of adult female cancer survivors in the U.S. in 1997 is 7.20 million (POPSIZE and POP CANCER); this estimate includes non-melanoma skin cancer. In this example, POPSIZE and POP CANCER are equal because the SUBPOPX statement restricts the analysis to those with cancer ever (CANEV =1, via CNKIND5 = 1 or 2), and the DENOM statement defines the denominator of the ratio as those with cancer ever (CANEV = 1). The estimated number of adult female breast cancer survivors in the U.S. (POP BRCA) in 1997 is 1.97 million. The proportion of adult female cancer survivors who have had breast cancer is  $(1,974,567)/(7,202,079) = 0.2742$ , with an estimated standard error of .0133. A 95% confidence interval on the percentage of adult female cancer survivors who have had breast cancer is  $27.42\% \pm 1.96 (1.33\%) = (24.80\%, 30.04\%)$ . The value 1.96 is from the standard normal distribution, which approximates the Student *t*-distribution with 339 df.