

Logistic (RLOGIST) Example #9

SUDAAN Statements and Results Illustrated

- Calculation of response rates and standard errors
- PREDSTAT
- RESPRATE
- SETENV
- NEST

Input Data Set(s): ELS.SAS7bdat

Example

Using data from the Education Longitudinal Study of 2002 (ELS:2002) second follow-up public-use file, model the probability of response in the base-year as a function of student race (FIRACE) and sex (BYSEX) and school region (BYREGION) and urbanicity (BYURBAN). Calculate the weighted response rate and associated standard error; overall and for each level of each explanatory variable.

Since the ELS:2002 public-use files do not include records for all base-year nonrespondents but only those base-year nonrespondents who responded in the first follow-up and since the analysis weights in the public-use files are adjusted for nonresponse, some leeway is required in order to carry out this example.

Since almost all base-year nonrespondents represented in the ELS:2002 second follow-up public-use file have a non-zero first follow-up cross-sectional weight (FIQWT), this weight will be used for the example. Ideally, the base-year design weight, not available in the public-use files, would be used instead of FIQWT.

Student race was adjusted for some respondents between base-year and first follow-up; either because no race information was available in the base-year (in the case of base-year nonrespondents who responded in the first follow-up) or because the original race classification was found to be in error. For this example, the first follow-up race variable (FIRACE) is used.

Solution

This example uses PROC RLOGIST (SAS-Callable SUDAAN) to model the probability of response in the base-year as a function of student characteristics (race and sex) and school characteristics (region and urbanicity). The data were extracted from the ELS:2002 second follow-up public-use file.

This example highlights the use of the PREDSTAT statement and the estimation weighted response rates and their standard errors.

This example was run in SAS-Callable SUDAAN, and the SAS program and *.LST files are provided.

SAS data step statements are used to create a binary variable to indicate base-year response status (1=base year respondent, 0=base-year nonrespondent) and to set negative values of the model covariates (FIRACE, BYREGION, BYURBAN, and BYSEX) to missing.

The CLASS statement tells SUDAAN to treat the listed variables as categorical. The NEST statement is used to specify the sampling strata and primary sampling unit variables STRAT_ID and PSU, respectively.

The SETENV statement is optional. They set up default formats for printed statistics and manipulate the printout to the needs of the user.

The MODEL statement is used to specify the variable that indicates response status (BYRESP) and the variables for which response rates will be calculated.

The WEIGHT statement specifies the weight variable to use for calculating R-indicators and propensity statistics.

The PREDSTAT statement is used to tell SUDAAN to calculate response rates and associated standard errors.

Exhibit 1. SAS-Callable SUDAAN Code

```
options ls=120 ps=68 pageno=1;
libname in "c:\ELS\ ";
options fmtsearch=(in);

data els;
  set in.els;
  /** Set up BY Response Status Indicator ***/
  if bysqstat=0 then byresp=0;else byresp=1;
  /** Since ELS:2002 variables use negative values (reserve codes) to indicate logical
  information; such as missing, nonrespondent, and legitimate skip, set reserve codes to missing
  **/
  array thevars{4} flrace byregion byurban bysex ;
  do i=1 to 4;
  if thevars{i}<0 then thevars{i}=.;
  end;
run;

proc rlogist data=els design=WR;
class flrace byregion byurban bysex ;
nest strat_id psu;
setenv decwidth=6;
model byresp=flrace byregion byurban bysex ;
weight flqwt;
PREDSTAT RESPRATE ;
run;
```

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      June 2012
Release 11.0.0-testing-221
```

```
DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a
With Replacement (WR) Design
Sample Weight: F1QWT
Stratification Variables(s): STRAT_ID
Primary Sampling Unit: PSU
```

Number of zero responses : 105
Number of non-zero responses : 14006

Independence parameters have converged in 7 iterations.

Number of observations read : 14930 Weighted count: 3466985
Number of observations skipped : 1267
(WEIGHT variable nonpositive)
Observations used in the analysis : 14111 Weighted count: 3233840
Denominator degrees of freedom : 390

Maximum number of estimable parameters for the model is 13

File ELS contains 751 Clusters
751 clusters were used to fit the model
Maximum cluster size is 48 records
Minimum cluster size is 2 records

Sample and Population Counts for Response Variable BYRESP
Based on observations used in the analysis
0: Sample Count 105 Population Count 23262
1: Sample Count 14006 Population Count 3210578

R-Square for dependent variable BYRESP (Cox & Snell, 1989): 0.003483

-2 * Normalized Log-Likelihood with Intercepts Only : 1204.05
-2 * Normalized Log-Likelihood Full Model : 1154.81
Approximate Chi-Square (-2 * Log-L Ratio) : 49.24
Degrees of Freedom : 12

Note: The approximate Chi-Square is not adjusted for clustering.
Refer to hypothesis test table for adjusted test.

Note from *Exhibit 2* that, under this example, there are 105 base-year nonrespondents and 14,006 base-year respondents. A total of 14,111 observations are used in the analysis.

Exhibit 3. Response Rates: Student Race

Variance Estimation Method: Taylor Series (WR)

SE Method: Robust (Binder, 1983)

Working Correlations: Independent

Link Function: Logit

Response variable BYRESP: BYRESP

by: Propensity and Weight Adjustment Statistics, F1 student^s race/ethnicity-composite.

Propensity and Weight Adjustment Statistics		F1 student^s race/ethnicity-composite				
		Total	Amer. Indian/- Alaska Native, non-Hispanic	Asian, Hawaii/- Pac. Islander, non-Hispanic	Black or African American, non-Hispanic	Hispanic, no race specified
Response Rate	Estimate	0.992807	0.992669	0.985787	0.988427	0.991043
	Standard Error	0.000960	0.007453	0.004256	0.002793	0.003921

Propensity and Weight Adjustment Statistics		F1 student^s race/ethnicity-composite		
		Hispanic, race specified	More than one race, non-Hispanic	White, non-Hispanic
Response Rate	Estimate	0.987402	0.994866	0.995010
	Standard Error	0.004257	0.002946	0.001080

The results show in the output (*Exhibit 3*, above) show, overall and for each level of student race (F1RACE), the response rate and standard error.

Notice that the response rates are all close to 1. There is some variation in response rates across the seven race/ethnicity groups. The response rate is highest for White, non-hispanics and lowest for Asian, Hawaiian/Pacific Islander, non-hispanics.

Exhibit 4. Response Rates: School Region

Variance Estimation Method: Taylor Series (WR)
 SE Method: Robust (Binder, 1983)
 Working Correlations: Independent
 Link Function: Logit
 Response variable BYRESP: BYRESP
 by: Propensity and Weight Adjustment Statistics, Geographic region of school.

		Geographic region of school				
Propensity and Weight Adjustment Statistics		Total	Northeast	Midwest	South	West
		Response Rate	Estimate	0.992807	0.994968	0.992259
	Standard Error	0.000960	0.001876	0.002278	0.001810	0.001342

The results show in the output (*Exhibit 4*, above) show, overall and for each level of school region, the response rate and standard error.

Notice that the response rates are all close to 1. There is some variation in response rates across the regions. The response rate is highest for students in schools in the West and lowest for students in schools in the South.

Exhibit 5. Response Rates: School Urbanicity

Variance Estimation Method: Taylor Series (WR)
 SE Method: Robust (Binder, 1983)
 Working Correlations: Independent
 Link Function: Logit
 Response variable BYRESP: BYRESP
 by: Propensity and Weight Adjustment Statistics, School urbanicity.

		School urbanicity			
Propensity and Weight Adjustment Statistics		Total	Urban	Suburban	Rural
		Response Rate	Estimate	0.992807	0.992890
	Standard Error	0.000960	0.001459	0.001575	0.001589

The results show in the output (*Exhibit 4*, above) show, overall and for each level of school urbanicity, the response rate and standard error.

Notice that the response rates are all close to 1. There is some variation in response rates between students in urban, suburban, and rural schools. The response rate is highest for students in rural schools and lowest for students in suburban schools.

Exhibit 6. Response Rates: Student Sex

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Logit
Response variable BYRESP: BYRESP
by: Propensity and Weight Adjustment Statistics, Sex-composite.

		Sex-composite		
Propensity and Weight Adjustment Statistics		Total	Male	Female
Response Rate	Estimate	0.992807	0.991018	0.994620
	Standard Error	0.000960	0.001349	0.001173

The results show in the output (*Exhibit 6*, above) show, overall and for each level of student sex, the response rate and standard error.

Notice that the response rates are all close to 1. There is some variation in response rates between males and females. The response rate is highest for Females.