Logistic (RLOGIST) Example #8

SUDAAN Statements and Results Illustrated

- Calculates R-indicator and propensity statistics
- PREDSTAT
- PSTD
- PVAR
- PMEAN
- PRSTD

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 (F1RACE) and sex (BYSEX) and school region (BYREGION) and urbanicity (BYURBAN). Calculate the R-indicator, propensity statistics, and standard errors; 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 (F1QWT), 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 F1QWT.

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 (F1RACE) 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, the estimation of R-indicators and their standard errors, the estimation of mean propensity, standard deviation of response propensities, variance of response propensities, and relative standard deviation of response propensities.

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 (F1RACE, 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 R-indicators and propensity statistics 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 R-indicators, the mean response propensity, the standard deviation and variance of response propensities, and the relative standard deviation of response propensities.

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} f1race byregion byurban bysex ;
do i=1 to 4;
if thevars{i}<0 then thevars{i}=.;
end;
run;

proc rlogist data=els design=WR;
class f1race byregion byurban bysex  ;
nest strat_id psu;
setenv decwidth=6;
model byresp=f1race byregion  byurban bysex  ;
weight f1qwt;
PREDSTAT RIND PSTD PVAR PMEAN PRSTD;
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.   R-Indicators and Response Propensity Statistics: 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. Islande- r,non- Hispanic	Black or African America- n, non- Hispanic	Hispani- c, no race specifi- ed
R-Indicator	Estimate	0.989004	0.990661	0.980606	0.989962	0.987657
	Standard Error	0.003048	0.009297	0.008402	0.003228	0.006363
Population Standard Deviation of Response Propensities	Estimate	0.005498	0.004669	0.009697	0.005019	0.006172
	Standard Error	0.001524	0.004649	0.004201	0.001614	0.003181
Population Variance of Response Propensities	Estimate	0.000030	0.000022	0.000094	0.000025	0.000038
	Standard Error	0.000017	0.000043	0.000081	0.000016	0.000039
Mean of Response Propensities	Estimate	0.992807	0.992669	0.985787	0.988427	0.991043
	Standard Error	0.000960	0.007453	0.004256	0.002793	0.003921
Relative Standard Deviation of Response Propensities	Estimate	0.005538	0.004704	0.009837	0.005078	0.006227
	Standard Error	0.001538	0.004717	0.004296	0.001642	0.003231

```
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.
```

| | | F1 student^s race/ethnicity-composite | | |
Propensity and Weight Adjustment Statistics		Hispanic, race specified	More than one race, non-Hispanic	White, non-Hispanic
R-Indicator	Estimate	0.983627	0.993845	0.995011
	Standard Error	0.007735	0.003789	0.001503
Population Standard Deviation of Response Propensities	Estimate	0.008186	0.003078	0.002494
	Standard Error	0.003867	0.001894	0.000752
Population Variance of Response Propensities	Estimate	0.000067	0.000009	0.000006
	Standard Error	0.000063	0.000012	0.000004
Mean of Response Propensities	Estimate	0.987402	0.994866	0.995010
	Standard Error	0.004257	0.002946	0.001080
Relative Standard Deviation of Response Propensities	Estimate	0.008291	0.003094	0.002507
	Standard Error	0.003948	0.001913	0.000757

The results show in the output (*Exhibit 3*, above) show, overall and for each level of student race (F1RACE), the r-indicator, the mean response propensity, standard deviation and variance of the response propensities, and the relative standard deviation of the response propensities. The standard error is also shown for each of these statistics.

Notice that the r-indicators and mean response propensities are close to 1; this occurs because there is a high overall response rate. There is some variation in r-indicators and mean response propensities across the seven race/ethnicity groups. The R-indicator and mean response propensity are highest for White, non-hispanics and lowest for Asian, Hawaiian/Pacific Islander, non-hispanics. Similarly, the population standard deviation, variance, and relative standard deviation of response propensities are lowest for White, non-hispanics and highest for Asian, Hawaiian/Pacific Islander, non-hispanics.

Exhibit 4.   R-Indicators and Response Propensity Statistics: 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.
```

Propensity and Weight Adjustment Statistics		Geographic region of school				
		Total	Northeast	Midwest	South	West
R-Indicator	Estimate	0.989004	0.993470	0.990288	0.987574	0.995824
	Standard Error	0.003048	0.002736	0.003572	0.004349	0.001874
Population Standard Deviation of Response Propensities	Estimate	0.005498	0.003265	0.004856	0.006213	0.002088
	Standard Error	0.001524	0.001368	0.001786	0.002175	0.000937
Population Variance of Response Propensities	Estimate	0.000030	0.000011	0.000024	0.000039	0.000004
	Standard Error	0.000017	0.000009	0.000017	0.000027	0.000004
Mean of Response Propensities	Estimate	0.992807	0.994968	0.992259	0.989554	0.996736
	Standard Error	0.000960	0.001876	0.002278	0.001810	0.001342
Relative Standard Deviation of Response Propensities	Estimate	0.005538	0.003282	0.004894	0.006279	0.002095
	Standard Error	0.001538	0.001379	0.001807	0.002205	0.000943

The results show in the output (*Exhibit 4*, above) show, overall and for each level of school region, the r-indicator, the mean response propensity, standard deviation and variance of the response propensities, and the relative standard deviation of the response propensities. The standard error is also shown for each of these statistics.

Notice that the r-indicators and mean response propensities are close to 1; this occurs because there is a high overall response rate. There is some variation in r-indicators and mean response propensities across the regions. The R-indicator and mean response propensity are highest for students in schools in the West and lowest for students in schools in the South. Similarly, the population standard deviation, variance, and relative standard deviation of response propensities are lowest for students in schools in the West and highest for students in schools in the South.

Exhibit 5.   R-Indicators and Response Propensity Statistics: 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
R-Indicator	Estimate	0.989004	0.989810	0.987848	0.992008
	Standard Error	0.003048	0.002771	0.003904	0.003313
Population Standard Deviation of Response Propensities	Estimate	0.005498	0.005095	0.006076	0.003996
	Standard Error	0.001524	0.001386	0.001952	0.001656
Population Variance of Response Propensities	Estimate	0.000030	0.000026	0.000037	0.000016
	Standard Error	0.000017	0.000014	0.000024	0.000013
Mean of Response Propensities	Estimate	0.992807	0.992890	0.992150	0.994348
	Standard Error	0.000960	0.001459	0.001575	0.001589
Relative Standard Deviation of Response Propensities	Estimate	0.005538	0.005131	0.006124	0.004019
	Standard Error	0.001538	0.001400	0.001973	0.001670

The results show in the output (**Exhibit 5**, above) show, overall and for each level of school urbanicity, the r-indicator, the mean response propensity, standard deviation and variance of the response propensities, and the relative standard deviation of the response propensities. The standard error is also shown for each of these statistics.

Notice that the r-indicators and mean response propensities are close to 1; this occurs because there is a high overall response rate. There is some variation in r-indicators and mean response propensities across the urbanicities. The R-indicator and mean response propensity are highest for students in rural schools and lowest for students in suburban schools. Similarly, the population standard deviation, variance, and relative standard deviation of response propensities are lowest for students rural schools and highest for students in suburban schools.

Exhibit 6.   R-Indicators and Response Propensity Statistics: 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.
```

Propensity and Weight Adjustment Statistics		Sex-composite		
		Total	Male	Female
R-Indicator	Estimate	0.989004	0.987390	0.992507
	Standard Error	0.003048	0.003818	0.002759
Population Standard Deviation of Response Propensities	Estimate	0.005498	0.006305	0.003747
	Standard Error	0.001524	0.001909	0.001379
Population Variance of Response Propensities	Estimate	0.000030	0.000040	0.000014
	Standard Error	0.000017	0.000024	0.000010
Mean of Response Propensities	Estimate	0.992807	0.991018	0.994620
	Standard Error	0.000960	0.001349	0.001173
Relative Standard	Estimate	0.005538	0.006362	0.003767
	Standard Error	0.001538	0.001930	0.001390

```
| Deviation of    |                       |           |           |           |
| Response        |                       |           |           |           |
| Propensities    |                       |           |           |           |
-------------------------------------------------------------------------
```

The results show in the output (***Exhibit 6***, above) show, overall and for each level of student sex, the r-indicator, the mean response propensity, standard deviation and variance of the response propensities, and the relative standard deviation of the response propensities. The standard error is also shown for each of these statistics.

Notice that the r-indicators and mean response propensities are close to 1; this occurs because there is a high overall response rate.   There is some variation in r-indicators and mean response propensities between males and females. The R-indicator and mean response propensity are highest for Females. Similarly, the population standard deviation, variance, and relative standard deviation of response propensities are lowest for Females.