

MULTILOG Example #3

SUDAAN Statements and Results Illustrated

- REFLEVEL
- CUMLOGIT option
- SETENV
- LEVELS
- WEIGHT

Input Data Set(s): IRONSUD.SSD

Example

Using data from the NHANES I and its Longitudinal Follow-up Study, model the effects of body iron stores, age group, and smoking status on follow-up cancer status. Compare the results using the default reference cells for categorical covariates against the same model formulated with the first level of each categorical variable defined as the reference level (using the REFLEVEL statement).

Solution

The following example comes from the NHANES I Study and its Longitudinal Follow-up Study conducted 10 years later.

In this analysis, we wish to determine whether follow-up cancer status (CANCER12, 1=yes, 2=no) is associated with a measure of body iron stores at the initial exam (B_TIBC, total iron-binding capacity), while adjusting for age group at initial exam (AGEGROUP, 1=20-49, 2=50+) and smoking status (SMOKE, 1=current, 2=former, 3=never, 4=unknown).

First, we supply results with the default reference cells, which is the last level of each categorical covariate that appears on the SUBGROUP statement (*i.e.*, SMOKE=4 and AGEGROUP=2).

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

Exhibit 1. SAS-Callable MULTLOG Code (Default Reference Cells)

```
LIBNAME IN V604 "C:\Program Files\SUDAAN\SUDAAN10\data";

PROC MULTLOG DATA=IN.IRONSUD FILETYPE=SAS DESIGN=WR DEFT2;
  NEST Q_STRATA PSU1;
  WEIGHT B_WTIRON;

  SUBGROUP CANCER12 AGEGROUP SMOKE;
  LEVELS 2 2 4;
  MODEL CANCER12 = B_TIBC AGEGROUP SMOKE / CUMLOGIT;

  SETENV COLSPCE=1 LABWIDTH=25 COLWIDTH=8 DECWIDTH=4;
  PRINT BETA="BETA" SEBETA="S.E." DEFT="DESIGN EFFECT" T_BETA="T:BETA=0"
        P_BETA="P-VALUE" DF WALDCHI WALDCHP / T_BETAfmt=F8.2 DEFTfmt=F6.2
        WALDCHIFMT=F8.2 DFFMT=F8.0;
  RTITLE "Default Reference Cell Model";
```

CANCER12 is the outcome variable in the model, while B_TIBC, AGEGROUP and SMOKE are covariates (*Exhibit 1*). In MULTLOG, the categorical response variable and all covariates that are to be modeled as categorical must appear on the SUBGROUP/LEVELS or CLASS statement.

The SETENV and PRINT statements are optional. The PRINT statement is used in this example to request individual statistics of interest, to change default labels for those statistics, and to specify a variety of formats for those printed statistics. Without the PRINT statement, default statistics are produced from each PRINT group, with default formats. The SETENV statement sets up default formats for printed statistics and further manipulates the printout to the needs of the user.

Exhibit 2. First Page of MULTLOG Output

```

                S U D A A N
      Software for the Statistical Analysis of Correlated Data
      Copyright      Research Triangle Institute      November 2011
                Release 11.0.0

DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method,
Assuming a With Replacement (WR) Design
      Sample Weight: B_WTIRON
      Stratification Variables(s): Q_STRATA
      Primary Sampling Unit: PSU1

Independence parameters have converged in 6 iterations

Number of observations read      :    3290      Weighted count: 40570323
Observations used in the analysis :    3290      Weighted count: 40570323
Denominator degrees of freedom   :         35

Maximum number of estimable parameters for the model is 6

File IN.IRONSUD contains 67 Clusters
67 clusters were used to fit the model
Maximum cluster size is 111 records
Minimum cluster size is 15 records

Sample and Population Counts for Response Variable CANCER12
Based on observations used in the analysis
1: Sample Count      232      Population Count  1745695
2: Sample Count      3058     Population Count  38824628

-2 * Normalized Log-Likelihood with Intercepts Only : 1167.64
-2 * Normalized Log-Likelihood Full Model           : 1026.95
Approximate Chi-Square (-2 * Log-L Ratio)           : 140.70
Degrees of Freedom                                  :          5

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

Exhibit 2 indicates that there are 3,290 records on the file, corresponding to 67 clusters, with a minimum and maximum cluster size of 15 and 111, respectively. There are no missing values in the data set and no SUBPOPN statement to subset the analysis, so all observations on the file are used in fitting the model. SUDAAN displays the weighted and unweighted frequency distributions of the response in the data and the number of iterations needed to estimate the regression coefficients.

Exhibit 3. Estimated Regression Coefficients (Default Reference Cells)

```

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Cumulative Logit
Response variable CANCER12: Cancer Status (1/2)

Default Reference Cell Model
-----
CANCER12 (cum-logit),
Independent Variables
and Effects
BETA      S.E.      DESIGN
EFFECT    T:BETA=0  P-VALUE
-----
CANCER12 (cum-logit)
Intercept 1          -0.8618    0.6605    0.94    -1.30    0.2004
TOTAL IRON BINDING
CAPACITY          -0.0024    0.0018    1.10    -1.29    0.2052
Age Cohort
20-49 yrs.        -2.2525    0.3343    1.89    -6.74    0.0000
50+ yrs.           0.0000    0.0000    .        .        .
Smoking Status
Current            -0.5858    0.2771    0.77    -2.11    0.0417
Former             -0.9418    0.2922    0.84    -3.22    0.0027
Never              -0.4998    0.2743    0.85    -1.82    0.0770
Unknown            0.0000    0.0000    .        .        .
-----

```

In this analysis (*Exhibit 3*), each smoking group is automatically compared to the unknown smoking status (SMOKE=4), which may not be very meaningful.

Exhibit 4. ANOVA Table

```

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Cumulative Logit
Response variable CANCER12: Cancer Status (1/2)

Default Reference Cell Model
-----
Contrast          Degrees
of Freedom        Wald      P-value
                  ChiSq     Wald
                  ChiSq     ChiSq
-----
OVERALL MODEL          6      708.28    0.0000
MODEL MINUS INTERCEPT 5      64.47    0.0000
B_TIBC                 1       1.67    0.1967
AGEGROUP               1      45.39    0.0000
SMOKE                  3      10.60    0.0141
-----

```

In the ANOVA table (*Exhibit 4*), we see that age group and smoking status are significantly associated with follow-up cancer status, but total iron-binding capacity is not ($p=0.1967$).

Using the REFLEVEL Statement

Next, using the REFLEVEL statement, we redefine the reference cells to be the first level of each categorical variable. Note that the only differences in the results are in the estimates of the regression

coefficients, where the expected value of the response for each level of the categorical covariate(s) is now compared to the user-specified first level instead of the last. The main effects tests remain unchanged.

Exhibit 5. SAS-Callable MULTLOG Code (REFLEVEL Statement)

```

PROC MULTLOG DATA="IN.IRONSUD" FILETYPE=SAS DESIGN=WR DEFT2;
NEST Q_STRATA PSU1;
WEIGHT B_WTIRON;

REFLEVEL AGEGROUP=1 SMOKE=1;
SUBGROUP CANCER12 AGEGROUP SMOKE;
LEVELS 2 2 4;
MODEL CANCER12 = B_TIBC AGEGROUP SMOKE / CUMLOGIT;

SETENV COLSPACE=1 LABWIDTH=25 COLWIDTH=8 DECWIDTH=4;
PRINT BETA="BETA" SEBETA="S.E." DEFT="DESIGN EFFECT" T_BETA="T:BETA=0"
      P_BETA="P-VALUE" DF WALDCHI WALDCHP / T_BETAfmt=F8.2 DEFTfmt=F6.2
      WALDCHIFMT=F8.2 DFFMT=F8.0;
RTITLE "Using the REFLEVEL Statement";

```

Exhibit 6. Estimated Regression Coefficients (REFLEVEL Statement)

```

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Cumulative Logit
Response variable CANCER12: Cancer Status (1/2)

Using the REFLEVEL Statement
-----
CANCER12 (cum-logit),
Independent Variables
and Effects          BETA      S.E.  DESIGN EFFECT  T:BETA=0  P-VALUE
-----
CANCER12 (cum-logit)
Intercept 1         -3.7002   0.6967   1.06    -5.31    0.0000
TOTAL IRON BINDING
CAPACITY            -0.0024   0.0018   1.10    -1.29    0.2052
Age Cohort
20-49 yrs.          0.0000   0.0000   .        .        .
50+ yrs.            2.2525   0.3343   1.89     6.74    0.0000
Smoking Status
Current              0.0000   0.0000   .        .        .
Former              -0.3560   0.2716   1.16    -1.31    0.1985
Never                0.0860   0.2500   1.26     0.34    0.7330
Unknown              0.5858   0.2771   0.77     2.11    0.0417
-----

```

Now each smoking group is compared to the current smokers (SMOKE=1), and we see immediately that current smokers are not significantly different from former smokers ($p=0.1985$), nor from those who have never smoked ($p=0.7330$). These results are displayed in *Exhibit 6*, above.

Exhibit 7. ANOVA Table

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Cumulative Logit
Response variable CANCER12: Cancer Status (1/2)

Using the REFLEVEL Statement

Contrast	Degrees of Freedom	Wald ChiSq	P-value Wald ChiSq
OVERALL MODEL	6	708.28	0.0000
MODEL MINUS INTERCEPT	5	64.47	0.0000
B_TIBC	1	1.67	0.1967
AGEGROUP	1	45.39	0.0000
SMOKE	3	10.60	0.0141

As noted earlier in this example, the tests of main effects are the same no matter which groups are designated as the reference cells.