SURVIVAL Example #1

SUDAAN Statements and Results Illustrated

- Accounting for intracluster correlation in survival analysis
- EVENT
- CLASS
- EFFECTS
- REFLEVEL

Input Data Set(s): IRONSUD.SSD

Example

Use data from the NHANES Longitudinal Follow-up Study to determine the effect of body iron stores on cancer onset time.

This example also highlights the new UNITS option to obtain multiple unit hazard ratios associated with continuous covariates.

Solution

Do body iron stores increase the risk of cancer in men? This question was answered using data from the National Health and Nutrition Examination Survey (NHANES I) and its epidemiologic follow-up.

In the main survey, a dietary questionnaire and medical exam provided information on transferrin saturation (expressed as a percentage) as an indicator of body iron stores:

$$\textit{Transferrin Saturation (\%)} = 100 * \left(\frac{\textbf{serum iron}}{\textbf{total iron - binding capacity}} \right)$$

During the epidemiologic follow-up, researchers recorded the *incidence* of cancer and *time* to cancer (in days) since initial exam. Censoring occurs at the time of follow-up interview if cancer incidence had not occurred by that time.

The data for this example consist of 3,290 men who

- had their total iron-binding capacity measured in NHANES I;
- had their cancer status determined in the follow-up survey; and
- were alive and cancer-free for at least 4 years after their biochemical measurements.

The explanatory variable of interest is a binary indicator (TRFSAT) for high transferrin saturation vs. normal levels ("high" transferrin saturation is defined as being in the highest quartile for this variable). In this analysis, we wish to determine whether *high transferrin saturation* is significantly associated with *time to cancer* in days (DAYFOLL is the response variable), after adjustment for age at initial examination (AGEXAM) and smoking status (SMOKE). People who did not develop cancer (CANCER1 = 0) by the follow-up interview are censored at that time.

Exhibit 1 below contains the results of statistical analyses in SUDAAN vs. SAS. Although both procedures estimate the same parameter vectors, the SAS results are all highly significant. This is due, in

part, to ignoring the intracluster correlation, and to the way SAS uses the analysis weights. SAS assumes there are over 40 million people in the sample (the sum of the weights). The default handling of ties (EFRON) is used in this example. In the models presented below, the reference cell for the Smoking Status variable is Unknown (default last level of Smoking Status).

Exhibit 1. Proportional Hazards Regression, NHANES Cancer Follow-Up Data

Covariate	Model-Fitting Method	β	S.E.	Design Effect[1]	Z	P-Value
High Transferrin Saturation	SUDAAN	0.1654	0.19684	1.16	0.84	.4065
	SAS (independence)	0.1654	0.00165		100.50	.0000
	SAS (normalized DF[2])	0.1654	0.18320		0.90	.3472
Age at Initial Exam	SUDAAN	0.0957	0.01006	1.55	9.51	.0000
	SAS (independence)	0.0957	0.00007		1315.50	.0000
	SAS (normalized DF[2])	0.0957	0.00808		11.84	.0000
Smoking Status - Current	SUDAAN	-0.6410	0.26283	0.77	-2.44	.0200
	SAS (independence)	-0.6410	0.00269		-237.90	.0000
	SAS (normalized DF[2])	-0.6410	0.29872		-2.14	.0324
Smoking Status - Former	SUDAAN	-0.9925	0.27338	0.79	-3.63	.0009
	SAS (independence)	-0.9925	0.00276		-359.30	.0000
	SAS (normalized DF[2])	-0.9925	0.30650		-3.23	.0012
Smoking Status - Never	SUDAAN	-0.3404	0.27457	0.90	-1.24	.2233
	SAS (independence)	-0.3404	0.00260		-130.90	.0000
	SAS (normalized DF[2])	-0.3404	0.28873		-1.18	.2380

[1] Design Effect $= \dfrac{V_{SUDAAN}}{V_{SAS(normalized\ DF)}}$

[2] Normalized DF: SAS SEs multiplied by a factor of $\sqrt{\dfrac{population\ size}{sample\ size}}$

Exhibit 2. SAS-Callable SUDAAN Code: SURVIVAL Procedure

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

proc format;
  value smoke 1="Current"
              2="Former"
              3="Never"
              4="Unknown";

PROC SURVIVAL DATA=in.IRONSUD FILETYPE=SAS DESIGN=WR;
  NEST Q_STRATA PSU1;
  WEIGHT B_WTIRON;

  EVENT CANCER1;
  CLASS SMOKE;
  REFLEVEL SMOKE=3;
  MODEL DAYFOLL = AGEXAM SMOKE TRFSAT;

  EFFECTS SMOKE=(1 -1 0 0) / exp NAME = "Current vs Former Smoke";
  EFFECTS AGEXAM / exp units=10 name="Age 10-yr increase";

  SETENV COLSPCE=1 TOPMGN=0 COLWIDTH=7 DECWIDTH=4 LABWIDTH=25;
  PRINT / betas=default tests=default
          t_betafmt=f6.2 deftfmt=f6.2 waldffmt=f6.2 dffmt=f7.0 ;

  SETENV COLSPCE=1 TOPMGN=0 COLWIDTH=5 DECWIDTH=3 LABWIDTH=25;
  PRINT / risk=default expcntrst=default
          hrfmt=f7.3 exp_cntrstfmt=f13.3 unitsfmt=f5.0;

  RFORMAT SMOKE SMOKE.;
  RLABEL SMOKE="Smoking Status";
  RTITLE "Relationship Between Body Iron Stores and Time to Cancer"
         "Adjusted for Age and Smoking Status";
```

Exhibit 2 contains the SAS-callable SUDAAN code for this example. The response variable is DAYFOLL, time to cancer (in days) since initial exam. Age at initial exam (AGEXAM) is modelled as a continuous covariate, while transferrin saturation level (TRFSAT) is binary (0=normal, 1=highest quartile) and therefore does not need to appear on a CLASS statement. Both of these variables are modeled with single regression coefficients. Smoking status (SMOKE) has four levels (current, former, never, unknown) and is modeled as a categorical covariate. The REFLEVEL statement is used here to change the reference cell of Smoking Status from Unknown to Never. The variable CANCER1 indicates a complete (cancer occurrence) vs. censored (no occurrence) time.

The EFFECTS statements will 1) test the effect of Current vs. Former Smokers, and 2) estimate the hazard ratio for Current to Former Smokers and the hazard ratio associated with a 10-year increase in Age.

Exhibit 3. First Page of SURVIVAL Output (*.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: B_WTIRON
    Stratification Variables(s): Q_STRATA
    Primary Sampling Unit: PSU1


NOTE: Using a default start time of -10000000000 for all records

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  5




Summary of Event Values
-----------------------------------------------------------
Cancer Status (0/1)          Frequency      Weighted Sum
-----------------------------------------------------------
Censored                      3058.000      38824628.000
Non-Censored                   232.000       1745695.000
-----------------------------------------------------------

SURVIVAL has converged to a solution in 6 iterations.

-2 * Normalized Log-Likelihood with Beta(s) = 0 :  4806.15
-2 * Normalized Log-Likelihood Full Model       :  4611.38
Approximate Chi-Square (-2 * Log-L Ratio)       :   194.77
Degrees of Freedom                              :        5
Approximate P-Value                             :     0.00

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

Exhibit 3 indicates that there are 232 cases where cancer was present vs. 3,058 subjects who were still cancer-free at follow-up. All 3,290 observations on the file are used in this analysis.

Exhibit 4. Frequencies for CLASS Variable SMOKE

```
Frequencies and Values for CLASS Variables
-----------------------------------
Smoking
  Status        Frequency      Value
-----------------------------------
Ordered
  Position:
  1                   881    Current
Ordered
  Position:
  2                   900     Former
Ordered
  Position:
  3                  1204      Never
Ordered
  Position:
  4                   305    Unknown
-----------------------------------
```

Exhibit 4 contains the frequencies and levels for all variables on the CLASS statement.

Exhibit 5 contains the estimated regression coefficients from the fitted model. Although the hazard for cancer onset is increased in the high transferrin saturation group (positive regression coefficient), body iron stores at the initial exam appear not to be significantly associated with follow-up time to cancer (p=0.4065). However, age at initial exam is positively associated with follow-up time to cancer (p=0.0000). Former smokers are the only group significantly different from those who never smoked (p=0.0083).

Exhibit 5. Estimated Regression Coefficients from SURVIVAL

```
Variance Estimation Method: Taylor Series (WR)
Dependent Variable: DAYFOLL: Time to Cancer (days)
Censoring Variable: CANCER1: Cancer Status (0/1)
Ties Handling: EFRON

Relationship Between Body Iron Stores and Time to Cancer
Adjusted for Age and Smoking Status

-------------------------------------------------------------------------------
Independent Variables and                      Lower     Upper
  Effects                                      95%       95%               P-value
                          Beta                 Limit     Limit     T-Test  T-Test
                          Coeff.    SE Beta    Beta      Beta      B=0     B=0
-------------------------------------------------------------------------------
Smoking Status
  Current                -0.3006    0.2640   -0.8364    0.2353    -1.14    0.2626
  Former                 -0.6520    0.2331   -1.1253   -0.1788    -2.80    0.0083
  Never                   0.0000    0.0000    0.0000    0.0000       .       .
  Unknown                 0.3404    0.2746   -0.2170    0.8978     1.24    0.2233
Age at Exam               0.0957    0.0101    0.0753    0.1161     9.51    0.0000
High Transferrin
  Saturation (0/1)        0.1654    0.1968   -0.2342    0.5650     0.84    0.4065
-------------------------------------------------------------------------------
```

Exhibit 6 contains the table of main effects tests and user-defined contrasts. Smoking status and age at initial exam are both significantly associated with follow-up time to cancer. The EFFECTS statement contrast indicates that current smokers are not significantly different from former smokers (a contrast not provided in the default regression coefficient output shown previously).

Exhibit 6. ANOVA Table and EFFECTS Statement Tests

```
Variance Estimation Method: Taylor Series (WR)
Dependent Variable: DAYFOLL: Time to Cancer (days)
Censoring Variable: CANCER1: Cancer Status (0/1)
Ties Handling: EFRON

Relationship Between Body Iron Stores and Time to Cancer
Adjusted for Age and Smoking Status
-----------------------------------------------------
Contrast                    Degrees
                            of                 P-value
                            Freedom   Wald F   Wald F
-----------------------------------------------------
OVERALL MODEL                     5    29.56    0.0000
SMOKE                             3     5.10    0.0050
AGEXAM                            1    90.43    0.0000
TRFSAT                            1     0.71    0.4065
Current vs Former Smoke           1     2.41    0.1299
-----------------------------------------------------
```

Exhibit 6 indicates that age at initial exam and smoking status significantly affect the time to cancer onset, but transferrin saturation level does not.

Exhibit 7 indicates that although cancer is observed to occur sooner in people with elevated transferrin saturation levels (hazards ratio = 1.18, implying an 18% increase in hazard), there is no statistically significant association, since the 95% confidence interval contains the null value of 1.0. The default hazard ratio of 1.100 for age at initial exam corresponds to a 10% increase in hazard for each 1-year increase in age at initial exam, and this is statistically significant (increased age at initial exam significantly associated with shortened time to cancer).

Exhibit 7. Default Hazard Ratios

Variance Estimation Method: Taylor Series (WR)
Dependent Variable: DAYFOLL: Time to Cancer (days)
Censoring Variable: CANCER1: Cancer Status (0/1)
Ties Handling: EFRON

Relationship Between Body Iron Stores and Time to Cancer
Adjusted for Age and Smoking Status

Independent Variables and Effects	Hazards Ratio	Lower 95% Limit	Upper 95% Limit
Smoking Status			
Current	0.740	0.433	1.265
Former	0.521	0.325	0.836
Never	1.000	1.000	1.000
Unknown	1.406	0.805	2.454
Age at Exam	1.100	1.078	1.123
High Transferrin Saturation (0/1)	1.180	0.791	1.759

Exhibit 8 contains the estimated user-specified hazard ratios obtained via the EXP and UNITS options on the EFFECTS statements. Current smokers have a 42% increase in hazard compared to former smokers (HR=1.421), but it is not statistically significant since 95% confidence limits contain the null value of 1.0. The hazard ratio of 2.604 associated with the 10-year increase in age corresponds to more than a doubling of hazard for each 10-year increase in age at initial exam, and this is statistically significant.

Exhibit 8. User-Specified Hazard Ratios

Variance Estimation Method: Taylor Series (WR)
Dependent Variable: DAYFOLL: Time to Cancer (days)
Censoring Variable: CANCER1: Cancer Status (0/1)
Ties Handling: EFRON

Relationship Between Body Iron Stores and Time to Cancer
Adjusted for Age and Smoking Status

Exponentiated Contrast	Units	EXP(Contrast)	Lower 95% Limit	Upper 95% Limit
Current vs Former Smoke	1	1.421	0.897	2.251
Age 10-yr increase	10	2.604	2.123	3.194