LOGLINK Example #1

SUDAAN Statements and Results Illustrated

- Log-linear regression modeling
- MODEL
- TEST
- SUBPOPN
- EFFECTS

Input Data Set(s): EPIL.SAS7bdat (Thall and Vail (1990))

Example

Use the Epileptic Seizure Data of Thall and Vail (1990) to fit a log-linear regression model to the number of epileptic seizures in each of 59 individuals in a treatment (progabide) and control (placebo) group.

Solution

In this example, the baseline interval length is not equal to the treatment interval lengths. In order to correct for this problem, define an offset or time-at-risk variable (McCullagh and Nelder, 1989), such that the log-linear model is as follows:

$$\log E(Y_{ij}) = \log\ t_{ij}\ +\ \mathbf{x}_{ij}^{/}\boldsymbol{\beta},$$

where Y_{ij} is the response count of events for the i-th patient (i=1,…,59) at the j-th interval (j=0,1,2,3,4, denoting baseline and the four 2-week intervals), $\mathbf{x}_{ij}$ is the vector of covariates, $\boldsymbol{\beta}$ describes the change in the log of the population average count per unit change in $\mathbf{x}_{ij}$, and t_{ij} is the offset variable. For the baseline period (j=0), $t_{ij} = 8$, and for each of the 2-week treatment period intervals (j=1,...,4), $t_{ij} = 2$. By definition, the regression coefficient for the offset variable is fixed at 1.

In this example, we model the data using the LOGLINK procedure, which is a generalized linear model with a log link function relating the expected value of the response to the covariates. To implement the GEE analysis, we specify the following:

- Exchangeable working correlations (R=EXCHANGEABLE).
- Robust variance estimator of Zeger and Liang (1986) (SEMETHOD=ZEGER).
- Patient ID as the cluster variable on the NEST statement (the file must also be sorted by ID).

Other options are later exercised to demonstrate the effects of ignoring the clustering, namely independent working correlations (R=INDEPENDENT, the default assumption) with a model-based variance estimator (SEMETHOD=MODEL).

The data file contains five records per patient, representing the single baseline and four treatment period values. Following is a listing of the data for two patients (one progabide patient, one placebo patient).

ID	TRT12	Interval	TIME12	OFFTIME	SEIZURES	SEIZ2W
101	1	1	1	2	11	11.00
101	1	2	1	2	14	14.00
101	1	3	1	2	9	9.00
101	1	4	1	2	8	8.00
101	1	0	2	8	76	19.00
	•					
	•					
	•					
104	2	1	1	2	5	5.00
104	2	2	1	2	3	3.00
104	2	3	1	2	3	3.00
104	2	4	1	2	3	3.00
104	2	0	2	8	11	2.75
	•					
	•					
	•					

Variable Definitions:
ID - Patient Identifier (this is the cluster on the SUDAAN NEST statement)

TRT12 (Treatment Group):
1 = Progabide
2 = Placebo

INTERVAL (Time Variable):
0 = Baseline
1-4 = Treatment Period (labeled "Post" in the output)

TIME12 (Time Variable):
1 = Treatment Period (labeled "Post" in the output)
2 = Baseline

OFFTIME (Offset variable)
= 2, if in treatment phase
= 8, if baseline

SEIZURES - number of seizures reporting during the interval (response variable in regression models)

SEIZ2W - 2-week seizure rates, for descriptive analyses only:
= SEIZURES, if in treatment phase
= SEIZURES / 4, if baseline

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

Exhibit 5. SAS-Callable SUDAAN Code (DESCRIPT Procedure)

```
libname in "c:\10winbetatest\exp\loglink";
options nocenter linesize=95 pagesize=60;

proc format;
 value trt12f 1="Progabide"
              2="Placebo";
 value time12f 1="Post"
               2="Baseline";

data one; set in.epil;
proc sort data=one; by id;

PROC DESCRIPT DATA=one FILETYPE=SAS NOMARG DEFT4;
NEST _ONE_ ID;
WEIGHT _ONE_;

SUBPOPN ID != 207 / NAME="ID 207 Deleted";
CLASS TRT12 TIME12;
VAR SEIZ2WK;
TABLES TRT12*TIME12;

SETENV COLWIDTH=8 DECWIDTH=2;
PRINT NSUM MEAN SEMEAN DEFFMEAN="Design Effect"/
      nsumfmt=f6.0 deffmeanfmt=f6.2 STYLE=NCHS;
rformat time12 time12f.;
rformat trt12 trt12f.;
rtitle "Thall and Vail Repeated Measures Data (1990)"
       "Descriptive Stats:  Average 2-Week Seizure Rates By Treatment and Time";
```

As in Diggle et al. (1994), we first present descriptive statistics on mean seizure rates per 2-week period, using the DESCRIPT procedure (***Exhibit 5*** through ***Exhibit 9***). The response variable SEIZ2WK is equal to the number of seizures in each of the 2-week treatment periods, and is equal to (the number of seizures/4) in the 8-week baseline period. In the regression analysis, we use the number of seizures as the response variable, and correct for unequal interval lengths using the OFFSET variable option on the MODEL statement.

The _ONE_ variable on the NEST and WEIGHT statements in ***Exhibit 5*** indicates that all the observations are from one stratum or block and that each observation is to receive a weight of one. Stratification and weighting are most often associated with analysis of sample surveys, but using the _ONE_ variable in the context of randomized experiments is a way to remove their effects when not needed.

Exhibit 6. First Page of SUDAAN Output for DESCRIPT (SAS *.LST File)

```
                                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: _ONE_
    Stratification Variables(s): _ONE_
    Primary Sampling Unit: ID


Number of observations read    :    295    Weighted count :      295
Observations in subpopulation  :    290    Weighted count :      290
Denominator degrees of freedom :     58
```

One patient was removed from the analysis, and hence there are 290 observations in the subpopulation used for analysis.

Exhibit 7. Frequencies for CLASS Variable GROUP

```
Frequencies and Values for CLASS Variables
by: Group.
---------------------------------------
Group           Frequency        Value
---------------------------------------
Ordered
  Position:
  1                   150    Progabide
Ordered
  Position:
  2                   140      Placebo
---------------------------------------
```

Exhibit 8. Frequencies for CLASS Variable TIME

```
Frequencies and Values for CLASS Variables
by: Time.
---------------------------------------
Time            Frequency       Value
---------------------------------------
Ordered
  Position:
  1                   232        Post
Ordered
  Position:
  2                    58    Baseline
---------------------------------------
```

There are more observations during the "Post" phase due to multiple observations (records) per patient during this time.

Exhibit 9. DESCRIPT Results

```
Variance Estimation Method: Taylor Series (WR)
For Subpopulation: ID 207 Deleted

Thall and Vail Repeated Measures Data (1990)
Descriptive Stats:  Average 2-Week Seizure Rates By Treatment and Time

for: Variable = Seizures (Per 2 Weeks).
-------------------------------------------------------------
Group
   Time                Sample                          Design
                       Size         Mean    SE Mean    Effect
-------------------------------------------------------------
Progabide
   Post                  120       5.71       0.91      2.92
   Baseline               30       6.91       0.79      0.98
Placebo
   Post                  112       8.60       1.64      2.80
   Baseline               28       7.70       1.22      0.98
-------------------------------------------------------------
```

The descriptive statistics results above indicate that average 2-week seizure rates decline slightly during the treatment period for the progabide group and increase slightly during the treatment period for the placebo group.

Note that the standard errors for the treatment phase (labeled "Post" in the output) have been appropriately inflated due to the multiple observations per patient. By default, SUDAAN uses a robust or between-cluster variance estimator to properly adjust for the repeated testing. The design effects show that the variance during the treatment phase had to be inflated almost threefold to properly account for the repeated testing. Also note that the design effect is near one for the baseline estimates since there is only one observation per patient at baseline.

GEE Regression Analysis Using Exchangeable Correlations

In the code below, we implement the GEE analysis by specifying exchangeable working correlations and using the robust variance estimator of Zeger and Liang (1986): R=EXCHANGEABLE and SEMETHOD=ZEGER.

Exhibit 10. SAS-Callable SUDAAN Code (GEE Exchangeable)

```
PROC LOGLINK DATA=one R=EXCHANGEABLE SEMETHOD=ZEGER;
NEST _ONE_ ID;
WEIGHT _ONE_;

SUBPOPN ID != 207 / NAME="ID 207 Deleted";
CLASS TRT12 TIME12;
MODEL SEIZURES = TIME12 TRT12 TRT12*TIME12 / LOGOFFSET=LTIME;
TEST WALDCHI;

EFFECTS TIME12=(1 -1) / TRT12=1 EXP NAME="Before Vs After: Progabide";
EFFECTS TIME12=(1 -1) / TRT12=2 EXP NAME="Before Vs After: Placebo";

SETENV COLWIDTH=7 DECWIDTH=4 LABWIDTH=26;
PRINT / betas=default tests=default idratio=default rhos=default
        T_BETAFMT=F6.2 DFFMT=F7.0 WALDCHIFMT=F6.2 IDRFMT=F9.4 LOWIDRFMT=F6.4
        UPIDRFMT=F6.4;

SETENV COLWIDTH=6 DECWIDTH=4 LABWIDTH=26;
PRINT / EXPCNTRST=DEFAULT EXP_CNTRSTFMT=F13.4;

rformat time12 time12f.;
rformat trt12 trt12f.;
RTITLE "Thall and Vail Repeated Measures Data (1990)"
       "GEE Regression Analysis";
```

On the MODEL statement in ***Exhibit 10***, we include an option for the offset term (LOGOFFSET=LTIME) defining the length of time during each interval. Model covariates include time (baseline vs. post), treatment (placebo vs. progabide), and their interaction. In this example, the primary focus is on the interaction between time and treatment regimen on the epileptic seizure rate. In other words, *does progabide reduce the rate of epileptic seizures?*

The EXP option on the EFFECTS statements in ***Exhibit 10*** will calculate the point estimate and confidence bounds for the IDR corresponding to the contrast that is requested (Before vs. After).

Exhibit 11. First Page of SUDAAN Output for GEE Exchangeable (SAS *.LST File)

```
                              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: _ONE_
    Stratification Variables(s): _ONE_
    Primary Sampling Unit: ID
    Cluster Identification Variables: _ONE_ ID


Independence parameters have converged in 7 iterations

Step 1 parameters have converged in 1 iteration.

Number of observations read       :    295    Weighted count:      295
Observations in subpopulation     :    290    Weighted count:      290
Observations used in the analysis :    290    Weighted count:      290
Denominator degrees of freedom    :     58


Maximum number of estimable parameters for the model is  4

File ONE contains   59 Clusters
  58 clusters were used to fit the model
Maximum cluster size is   5 records
Minimum cluster size is   5 records

Weighted mean response is 11.513793
```

Exhibit 12. Regression Coefficients for GEE Exchangeable

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Zeger-Liang, 1986)
Working Correlations: Exchangeable
Link Function: Log
Response variable SEIZURES: SEIZURES
LOG Offset variable LTIME: LTIME
For Subpopulation: ID 207 Deleted

Thall and Vail Repeated Measures Data (1990)
GEE Regression Analysis
```

Independent Variables and Effects	Beta Coeff.	SE Beta	Lower 95% Limit Beta	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0
Intercept	1.3476	0.1587	1.0299	1.6653	8.49	0.0000
Group						
Progabide	-0.1080	0.1953	-0.4990	0.2830	-0.55	0.5824
Placebo	0.0000	0.0000	0.0000	0.0000	.	.
Time						
Post	0.1108	0.1171	-0.1236	0.3452	0.95	0.3480
Baseline	0.0000	0.0000	0.0000	0.0000	.	.
Group, Time						
Progabide, Post	-0.3016	0.1727	-0.6472	0.0440	-1.75	0.0860
Progabide, Baseline	0.0000	0.0000	0.0000	0.0000	.	.
Placebo, Post	0.0000	0.0000	0.0000	0.0000	.	.
Placebo, Baseline	0.0000	0.0000	0.0000	0.0000	.	.

The coefficient of primary interest, the interaction parameter, is estimated to be -0.3016 (see ***Exhibit 12***), and it is only marginally significant (p=0.0860, based on a t-test with 58 df). Because we specified SEMETHOD=ZEGER on the PROC statement, the estimated standard errors of the regression coefficients are computed using the robust variance estimator of Zeger and Liang (1986). A model-based variance estimator may also have been valid here.

Although only marginally significant (p=0.0860), the negative regression coefficient for the interaction term in the model indicates a greater reduction in seizure counts for the progabide vs. control group during the treatment period vs. baseline. This is also reflected in the average 2-week seizure rates from the DESCRIPT procedure. For the control group, the seizure rate goes up in the treatment period, and for the progabide group, the seizure rate goes down during the treatment period.

The EFFECTS statements test the difference between baseline and treatment period separately for each treatment group. ***Exhibit 13*** indicates that the interaction term is marginally significant (p=0.0807, based on a Wald chi-square test). However, neither group significantly changed value from the baseline vs. treatment phase (p=0.1327 for the progabide group, p=0.3440 for the control group).

Exhibit 13. ANOVA Table for GEE Exchangeable

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Zeger-Liang, 1986)
Working Correlations: Exchangeable
Link Function: Log
Response variable SEIZURES: SEIZURES
LOG Offset variable LTIME: LTIME
For Subpopulation: ID 207 Deleted

Thall and Vail Repeated Measures Data (1990)
GEE Regression Analysis
```

Contrast	Degrees of Freedom	Wald ChiSq	P-value Wald ChiSq
OVERALL MODEL	4	192.80	0.0000
MODEL MINUS INTERCEPT	3	3.67	0.2991
INTERCEPT	.	.	.
TRT12	.	.	.
TIME12	.	.	.
TRT12 * TIME12	1	3.05	0.0807
Before Vs After: Progabide	1	2.26	0.1327
Before Vs After: Placebo	1	0.90	0.3440

Exhibit 14 indicates that the estimated exchangeable correlation among the five observations per patient is 0.5983.

Exhibit 14. Intracluster Correlation for GEE Exchangeable

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Zeger-Liang, 1986)
Working Correlations: Exchangeable
Link Function: Log
Response variable SEIZURES: SEIZURES
LOG Offset variable LTIME: LTIME
For Subpopulation: ID 207 Deleted

Thall and Vail Repeated Measures Data (1990)
GEE Regression Analysis

by: SEIZURES.
```

SEIZURES	Correlation Matrix
1	0.5983

Each regression coefficient describes the change in the log of the population average count per unit change in the covariate *x*. Exponentiating the estimated regression coefficient, $\exp(\beta)$, yields the estimated incidence density ratio (IDR). The IDR is defined to be the ratio of average seizure rates per unit interval of time for each one-unit increase in the model covariate.

In the interaction model, the default IDR table in ***Exhibit 15*** tells us that the IDR associated with the time effect (post vs. baseline) = 1.1172 for the placebo group. This represents an 11.7% increase in seizure rate from baseline to post, although the confidence limits contain the null value of 1.0 (no difference between baseline and post). The results for the progabide group are not revealed in the default IDR table, since progabide is not a reference cell in the interaction model. The EXP option on the EFFECTS statement is more useful for obtaining the IDRs of interest in interaction models (see ***Exhibit 16***).

Exhibit 15. Default Incidence Density Ratios for GEE Exchangeable

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Zeger-Liang, 1986)
Working Correlations: Exchangeable
Link Function: Log
Response variable SEIZURES: SEIZURES
LOG Offset variable LTIME: LTIME
For Subpopulation: ID 207 Deleted

Thall and Vail Repeated Measures Data (1990)
GEE Regression Analysis
```

Independent Variables and Effects	Incidence Density Ratio	Lower 95% Limit IDR	Upper 95% Limit IDR
Intercept	3.8482	2.8008	5.2872
Group			
Progabide	0.8976	0.6071	1.3271
Placebo	1.0000	1.0000	1.0000
Time			
Post	1.1172	0.8837	1.4123
Baseline	1.0000	1.0000	1.0000
Group, Time			
Progabide, Post	0.7396	0.5235	1.0450
Progabide, Baseline	1.0000	1.0000	1.0000
Placebo, Post	1.0000	1.0000	1.0000
Placebo, Baseline	1.0000	1.0000	1.0000

For interaction models, it is generally necessary to obtain the IDRs of interest using the EXP option on the EFFECTS statement. ***Exhibit 16*** contains the IDRs for both the Progabide and Placebo groups. For the progabide group (first row), the IDR for the pre vs. post comparison is *exp*(0.1108-0.3016)=0.8263, which indicates that the number of seizures has dropped by about 18%. However, as stated earlier, this is not a statistically significant decrease, and the confidence interval for the IDR contains the null value of 1.0. The second row is the IDR for the placebo group. The value of 1.1172 indicates about an 11% increase in the number of seizures; this is also not a statistically significant increase. Note this result was also indicated in the default IDR table.

Although the confidence intervals still contain the null value of 1.0, the IDR results describe the increase in seizure rates for the placebo group and a decrease for the progabide group.

Exhibit 16. User-Specified Incidence Density Ratios for GEE Exchangeable

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Zeger-Liang, 1986)
Working Correlations: Exchangeable
Link Function: Log
Response variable SEIZURES: SEIZURES
LOG Offset variable LTIME: LTIME
For Subpopulation: ID 207 Deleted

Thall and Vail Repeated Measures Data (1990)
GEE Regression Analysis

------------------------------------------------------------
Contrast                                    Lower    Upper
                                            95%      95%
                            EXP(Contrast)   Limit    Limit
------------------------------------------------------------
Before Vs After: Progabide          0.8263   0.6409   1.0653
Before Vs After: Placebo            1.1172   0.8837   1.4123
------------------------------------------------------------
```

Ordinary Regression Analysis Assuming Independence

In the code below (***Exhibit 17***), we conduct an ordinary log-linear regression analysis, incorrectly assuming independence among the repeated measures.

The R=INDEPENDENT (default working correlation) and SEMETHOD=MODEL options in ***Exhibit 17*** yield an ordinary log-linear regression analysis for count data that inappropriately assumes independence among the repeated measurements. We present these results only for comparison to the previous analysis, which properly accounts for the repeated testing.

Exhibit 17. SAS-Callable SUDAAN Code (Independence Assumption)

```
PROC LOGLINK DATA=one deft4 R=INDEPENDENT SEMETHOD=MODEL;
NEST _ONE_ ID;
WEIGHT _ONE_;

SUBPOPN ID != 207/NAME="ID 207 Deleted";
CLASS TRT12 TIME12;
MODEL SEIZURES = TIME12 TRT12 TRT12*TIME12 / LOGOFFSET=LTIME;
TEST WALDCHI;

EFFECTS TIME12=(1 -1) / TRT12=1 EXP NAME="Before Vs After: Progabide";
EFFECTS TIME12=(1 -1) / TRT12=2 EXP NAME="Before Vs After: Placebo";

SETENV COLWIDTH=7 DECWIDTH=4 LABWIDTH=26;
PRINT / betas=default tests=default idratio=default T_BETAFMT=F6.2 DFFMT=F7.0
        WALDCHIFMT=F7.2 IDRFMT=F9.4 LOWIDRFMT=F6.4 UPIDRFMT=F6.4;

SETENV COLWIDTH=6 DECWIDTH=4 LABWIDTH=26;
PRINT / EXPCNTRST=DEFAULT EXP_CNTRSTFMT=F13.4;

rformat time12 time12f.;
rformat trt12 trt12f.;
RTITLE "Thall and Vail Repeated Measures Data (1990)"
       "Regression Analysis Assuming Independence";
```

Exhibit 18. First Page of SUDAAN Output Assuming Independence (SAS *.LST File)

```
                                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: _ONE_
    Stratification Variables(s): _ONE_
    Primary Sampling Unit: ID


Independence parameters have converged in 7 iterations

Number of observations read       :    295    Weighted count:      295
Observations in subpopulation     :    290    Weighted count:      290
Observations used in the analysis :    290    Weighted count:      290
Denominator degrees of freedom    :     58


Maximum number of estimable parameters for the model is  4

File ONE contains   59 Clusters
  58 clusters were used to fit the model
Maximum cluster size is   5 records
Minimum cluster size is   5 records
Weighted mean response is 11.513793
```

Exhibit 19. Regression Coefficients Assuming Independence

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Model-Based (Naive)
Working Correlations: Independent
Link Function: Log
Response variable SEIZURES: SEIZURES
LOG Offset variable LTIME: LTIME
For Subpopulation: ID 207 Deleted

Thall and Vail Repeated Measures Data (1990)
Regression Analysis Assuming Independence
```

Independent Variables and Effects	Beta Coeff.	SE Beta	Lower 95% Limit Beta	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0
Intercept	1.3476	0.0341	1.2794	1.4158	39.57	0.0000
Group						
Progabide	-0.1080	0.0486	-0.2054	-0.0107	-2.22	0.0303
Placebo	0.0000	0.0000	0.0000	0.0000	.	.
Time						
Post	0.1108	0.0469	0.0169	0.2047	2.36	0.0215
Baseline	0.0000	0.0000	0.0000	0.0000	.	.
Group, Time						
Progabide, Post	-0.3016	0.0697	-0.4412	-0.1620	-4.32	0.0001
Progabide, Baseline	0.0000	0.0000	0.0000	0.0000	.	.
Placebo, Post	0.0000	0.0000	0.0000	0.0000	.	.
Placebo, Baseline	0.0000	0.0000	0.0000	0.0000	.	.

Note that the naïve standard errors assuming independence (***Exhibit 19***) are (inappropriately) much smaller than those using the robust variance estimator in the exchangeable analysis (***Exhibit 12***). As a result, the model effects are highly significant under independence, yielding invalid inference.

Exhibit 20. ANOVA Table Under Independence

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Model-Based (Naive)
Working Correlations: Independent
Link Function: Log
Response variable SEIZURES: SEIZURES
LOG Offset variable LTIME: LTIME
For Subpopulation: ID 207 Deleted

Thall and Vail Repeated Measures Data (1990)
Regression Analysis Assuming Independence

-----------------------------------------------------------
Contrast                     Degrees             P-value
                             of        Wald      Wald
                             Freedom   ChiSq     ChiSq
-----------------------------------------------------------
OVERALL MODEL                      4   5640.96    0.0000
MODEL MINUS INTERCEPT              3     72.13    0.0000
INTERCEPT                          .       .       .
TRT12                              .       .       .
TIME12                             .       .       .
TRT12 * TIME12                     1     18.70    0.0000
Before Vs After: Progabide         1     13.65    0.0002
Before Vs After: Placebo           1      5.58    0.0181
-----------------------------------------------------------
```

Assuming independence, the interaction and both user-specified contrasts are inappropriately statistically significant (see ***Exhibit 20***, above).

The incidence density ratio (IDR) output computed naïvely assuming independence is not displayed here. The 95% confidence intervals around the IDR estimates are tighter and statistically significant, due again to the (inappropriately) smaller standard errors computed under the naive assumption of independence.