REGRESS Example #2

SUDAAN Statements and Results Illustrated

- GEE linear regression
- Delete-1 Jackknife variance estimation
- Binder robust variance estimator
- TEST
- CONDMARG

Input Data Set(s): BORIC.SSD

Example

Teratology Experiment: Clustered Continuous Data. This example demonstrates the GEE (Zeger and Liang, 1986; Liang and Zeger, 1986) and Delete-1 jackknife model-fitting techniques for linear regression in the context of a preclinical teratology experiment.

This example also highlights the new confidence limits for predicted and conditional marginals introduced in SUDAAN 11.0.

Solution

The data for this example represent fetal body weight in rats after daily administration of boric acid (0, 0.025, 0.05, 0.075, 0.1, or 0.2% in feed) to the dam during gestation. The experiments included a total of 164 litters (average of 27 litters per group) and anywhere from two to 14 fetuses per litter (1,302 fetuses total). ***Exhibit 1*** shows the structure of the data.

Exhibit 1. Structure of the Fetal Body Weight Data

Dose Group *1 = Control* *6 = High Dose*	*Litter ID*	*Fetus ID*	*Y = fetal body weight (gms)*
1	1	1	3.56
1	1	2	3.20
1	1	3	4.14
1	2	1	2.99
1	2	2	3.21
6	10	1	2.11
6	10	2	3.43
6	20	1	4.88
6	20	2	3.10
6	30	1	2.67

N = 1,302 records on the file (1,302 fetuses clustered within 164 litters)

In this example, the observations on fetuses are clustered within litters. The design effect measures the inflation (or deflation) in variance of a sample statistic due to intracluster correlation beyond that

expected if the data were independent. It is estimated as the ratio of the cluster sample variance obtained through GEE or jackknife vs. independence. Design effects in this study ranged from 3 to 6, reflecting high intralitter correlations.

To implement the GEE linear regression methods in SUDAAN, we first estimated the model parameters under ordinary least squares and computed a robust variance estimate. This is the GEE linear model with independent working correlations (which we refer to as GEE-independent). The Wald chi-square test was used to evaluate the null hypothesis of no dose-related effect. For comparison, the same linear model was also fit using GEE linear regression under exchangeable intralitter correlations; using linear regression with Delete-1 jackknife variance estimation; and using standard statistical software under the assumption of independent observations. See ***Exhibit 2*** for results.

The results under the GEE-exchangeable, GEE-independent, and jackknife approaches were essentially the same. For comparing the high dose to control in the linear model, the GEE-exchangeable approach yielded a *Z*-statistic of -7.40, compared to a GEE-independent *Z*-statistic of -8.11 and a jackknife Z-statistic of -7.83. The observed design effect for the high dose vs. control regression parameter was over 3.0 for these data, reflecting substantial intralitter correlations (estimated to be 0.5056 for these data).

Exhibit 2. Linear Regression for Boric Acid Data: Exposed vs. Control

Contrast	Model-Fitting Method	β	S.E.	Z	P-Value
0.025% vs. Control	GEE (indep)	-0.0611	0.0596	-1.03	0.3067
	GEE (exch corr)	-0.0509	0.0622	-0.82	0.4148
	Jackknife	-0.0611	0.0616	-0.99	0.3227
	Independence	-0.0611	0.0332	-1.84	0.0676
0.050% vs. Control	GEE (indep)	-0.0789	0.0724	-1.09	0.2777
	GEE (exch corr)	-0.0656	0.0738	-0.89	0.3753
	Jackknife	-0.0789	0.0752	-1.05	0.2955
	Independence	-0.0789	0.0330	-2.39	0.0180
0.075% vs. Control	GEE (indep)	-0.1219	0.0740	-1.65	0.1016
	GEE (exch corr)	-0.1363	0.0790	-1.73	0.0864
	Jackknife	-0.1219	0.0765	-1.59	0.1128
	Independence	-0.1219	0.0327	-3.73	0.0003
0.10% vs. Control	GEE (indep)	-0.2062	0.0627	-3.29	0.0012
	GEE (exch corr)	-0.2409	0.0680	-3.54	0.0005
	Jackknife	-0.2062	0.0648	-3.18	0.0018
	Independence	-0.2062	0.0323	-6.39	0.0000
0.20% vs. Control	GEE (indep)	-0.4883	0.0602	-8.11	0.0000
	GEE (exch corr)	-0.4822	0.0651	-7.40	0.0000
	Jackknife	-0.4883	0.0624	-7.83	0.0000
	Independence	-0.4883	0.0336	-14.54	0.0000

Sources:

GEE (independent):	SUDAAN REGRESS Procedure
GEE (exchangeable):	SUDAAN REGRESS Procedure
Delete-1 Jackknife:	SUDAAN REGRESS Procedure
Independence:	Standard Packages

Naively ignoring the clustering of the design in both parameter and variance estimation yields significant reductions ($p<0.05$) in body weights in dose groups as low as .05% (second lowest dose group) and marginally significant reductions ($p=0.06$) in the lowest dose group, while all three alternative approaches (GEE-independent, GEE-exchangeable, and jackknife) only detect significant reductions in the two

highest dose groups. Therefore, if we ignore intracluster correlations for cluster-level covariates (dose group in this study), we run the risk of detecting false-positive results.

Descriptive Statistics

Here we present the average fetal body weight in each dose group, along with their estimated standard errors (using a robust variance estimate to adjust for clustering) and design effects. These design effects were in the range of 3 to 6, reflecting more than a tripling in the variance of the estimated means under the clustered design. In the program code (***Exhibit 2-2***), the DAMID variable represents the cluster on the NEST statement. ***Exhibit 2-6*** shows that fetal body weight is reduced in the two highest dose groups compared to the control.

Exhibit 3. SAS-Callable Sudaan Code: DESCRIPT

```
libname in "c:\11winbetatest\CIs for Marginals\REGRESS\Boric";
options linesize=95 pagesize=60;

proc format;
  value dose 1="1 = Control"
             2="2 = 0.025%"
             3="3 = 0.05%"
               4="4 = 0.075%"
               5="5 = 0.1%"
               6="6 = 0.2%";

PROC DESCRIPT DATA=in.boric FILETYPE=SAS NOMARG DESIGN=WR;
  NEST _ONE_ DAMID;
  WEIGHT _ONE_;
  VAR BW;
  CLASS DOSEGRP;
  SETENV LABWIDTH=25 COLWIDTH=8 DECWIDTH=4 ;
  PRINT NSUM="SAMPLE   SIZE"
        MEAN="MEAN"
        SEMEAN="SE"
        DEFFMEAN="DESIGN EFFECT" /
        STYLE=NCHS NSUMFMT=F6.0 deffmeanfmt=f6.2;
  RFORMAT dosegrp dose.;
  RTITLE "Fetal Body Weight in a Teratology Study";
```

Exhibit 4. First Page of DESCRIPT Output (*.lst file)

```
                                S U D A A N
           Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute     February 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: DAMID

Number of observations read    :   1302    Weighted count :     1302
Denominator degrees of freedom :    163
```

Exhibit 4 indicates that DESCRIPT read 1,302 records, and there are 163 DDF (164 litters – 1 stratum).

Exhibit 5. Frequencies and Values for CLASS Variable DOSEGRP

```
Frequencies and Values for CLASS Variables
by: DOSEGRP.
----------------------------------------
DOSEGRP          Frequency          Value
----------------------------------------
Ordered
  Position:
  1                    217    1 = Control
Ordered
  Position:
  2                    210     2 = 0.025%
Ordered
  Position:
  3                    215      3 = 0.05%
Ordered
  Position:
  4                    223     4 = 0.075%
Ordered
  Position:
  5                    236       5 = 0.1%
Ordered
  Position:
  6                    201       6 = 0.2%
----------------------------------------
```

Exhibit 5 contains the frequencies for CLASS variable DOSEGRP.

Exhibit 6. Mean Body Weight in Each Dose Group

```
Variance Estimation Method: Taylor Series (WR)

Fetal Body Weight in a Teratology Study
-------------------------------------------------------------------
Variable
   DOSEGRP                   SAMPLE                          DESIGN
                               SIZE       MEAN          SE   EFFECT
-------------------------------------------------------------------
Fetal Body Weight
   1 = Control                  217     3.4979     0.0406     3.22
   2 = 0.025%                   210     3.4367     0.0436     3.55
   3 = 0.05%                    215     3.4190     0.0600     5.41
   4 = 0.075%                   223     3.3760     0.0619     6.05
   5 = 0.1%                     236     3.2916     0.0477     4.80
   6 = 0.2%                     201     3.0096     0.0444     4.73
-------------------------------------------------------------------
```

Exhibit 6 shows that fetal body weight is reduced in the two highest dose groups compared to the control. Standard errors are adjusted for the clustered design.

REGRESS Runs for Teratology Data

The REGRESS procedure fits a linear regression model to the fetal body weights, with dose group (six-level categorical variable, from 1=*control* to 6=*high dose*) as the only predictor. The DAMID variable remains as the cluster on the NEST statement. We use the REFLEVEL statement to change the reference level for DOSEGRP from the default last level (*high dose*) to the first (*control*), so that the regression parameters will be comparing each treatment group to the control. We also request the conditional marginal means on the CONDMARG statement (otherwise known as least squares means for linear regression) within each dose group. In this case, the conditional marginal means will be equal to the raw means in each dose group, since DOSEGRP is the only covariate in the model.

We work with the R and SEMETHOD options to obtain the different GEE and jackknife variance methods, each described below.

GEE Under Independent Working Correlations

We specify ***R=Independent*** to estimate the model via GEE under independent "working" correlations. By default, SUDAAN will apply a between-cluster variance estimator (Binder, 1983). For linear regression models, the robust variance estimator of Binder (1983) is equivalent to that of Zeger and Liang (1986). The *R=Independent* option is also default, so technically we didn't need to include that either.

Exhibit 7 contains the SAS-Callable REGRESS code for GEE under working independence.

Exhibit 7.SAS-Callable SUDAAN Code: GEE independent

```
PROC REGRESS DATA=in.boric FILETYPE=SAS R=INDEPENDENT;
  NEST _ONE_ DAMID;
  WEIGHT _ONE_;

  REFLEVEL DOSEGRP = 1;
  CLASS DOSEGRP;
  MODEL BW = DOSEGRP;
  TEST WALDCHI;
  CONDMARG DOSEGRP;

  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" / TESTS=default t_betafmt=F8.2 deftfmt=F6.2 dffmt=F7.0
        waldchifmt=F10.2 waldchpfmt=f7.4;

  SETENV LABWIDTH=20 decwidth=4 colwidth=6;
  PRINT / COND_MRG=default condmrgfmt=f11.4 t_cndmrgfmt=f8.2 p_cndmrgfmt=f7.4;

  RFORMAT dosegrp dose.;
  RTITLE "Treatment Effect on Fetal Body Weight in a Teratology Experiment";
```

Exhibit 8. First Page of REGRESS Output: GEE independent

```
                                S U D A A N
           Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute     February 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: DAMID


Number of observations read       :   1302    Weighted count:     1302
Observations used in the analysis :   1302    Weighted count:     1302
Denominator degrees of freedom    :    163


Maximum number of estimable parameters for the model is  6

File IN.BORIC contains  164 Clusters
 164 clusters were used to fit the model
Maximum cluster size is  14 records
Minimum cluster size is   2 records

Weighted mean response is 3.341329

Multiple R-Square for the dependent variable BW: 0.171089
```

Exhibit 8 is the first page of printed output. REGRESS reports that there are 164 clusters, 1,302 fetuses (records on the file), and a minimum and maximum cluster size of 2 and 14, respectively.

Exhibit 9 contains the estimated regression coefficients under working independence. By default, SUDAAN uses the robust variance estimator of Binder (1983), which appropriately corrects for intracluster correlation and yields valid results. We see that the two highest dose groups have significantly lower body weights than controls (p=0.0012 and 0.0000 for the 0.1% and 0.2% groups vs. controls, respectively). Again, design effects for regression coefficients in the range of 3 to 4 indicate more than a tripling in the variance of the estimated regression coefficients under the clustered design. The SUDAAN standard errors appropriately reflect this increase.

Exhibit 9. Linear Regression Coefficients: GEE independent

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
------------------------------------------------------------------------------
Independent Variables and
  Effects                                          Design
                                  BETA       S.E.  Effect   T:BETA=0    P-VALUE
------------------------------------------------------------------------------
Intercept                       3.4979     0.0406     3.04      86.10     0.0000
DOSEGRP
  1 = Control                   0.0000     0.0000      .          .         .
  2 = 0.025%                   -0.0611     0.0596     3.22      -1.03     0.3067
  3 = 0.05%                    -0.0789     0.0724     4.81      -1.09     0.2777
  4 = 0.075%                   -0.1219     0.0740     5.12      -1.65     0.1016
  5 = 0.1%                     -0.2062     0.0627     3.77      -3.29     0.0012
  6 = 0.2%                     -0.4883     0.0602     3.21      -8.11     0.0000
------------------------------------------------------------------------------
```

The main effects tests in ***Exhibit 10*** indicate that the overall effect of treatment (with 5 degrees of freedom) is statistically significant, after adjusting for clustering (p=0.0000).

Exhibit 10. ANOVA Table: GEE independent

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
------------------------------------------------------------
Contrast                    Degrees                 P-value
                            of                      Wald
                            Freedom   Wald ChiSq    ChiSq
------------------------------------------------------------
OVERALL MODEL                     6     29189.67     0.0000
MODEL MINUS INTERCEPT             5        78.38     0.0000
INTERCEPT                         .           .          .
DOSEGRP                           5        78.38     0.0000
------------------------------------------------------------
```

Exhibit 11. Conditional Marginal Means: GEE independent

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-------------------------------------------------------------------------------
Conditional Marginal
  #1                                        Lower    Upper
                      Conditional           95%      95%
                      Marginal          SE  Limit    Limit    T:Marg=0   P-value
-------------------------------------------------------------------------------
DOSEGRP
  1 = Control              3.4979   0.0406   3.4176   3.5781      86.10    0.0000
  2 = 0.025%               3.4367   0.0436   3.3506   3.5229      78.80    0.0000
  3 = 0.05%                3.4190   0.0600   3.3006   3.5374      57.01    0.0000
  4 = 0.075%               3.3760   0.0619   3.2537   3.4982      54.54    0.0000
  5 = 0.1%                 3.2916   0.0477   3.1974   3.3859      68.95    0.0000
  6 = 0.2%                 3.0096   0.0444   2.9218   3.0973      67.73    0.0000
-------------------------------------------------------------------------------
```

The conditional marginals in ***Exhibit 11*** are identical to the raw means and standard errors presented in ***Exhibit 6***, since there is only one covariate in the model. Body weights appear to be reduced in the two highest dose groups vs. control. Confidence limits in the two highest dose groups do not overlap with that of controls, indicating statistically significant differences among these groups.

Independence via Model-Based (Naive) Variance Estimates

Exhibit 12 contains the code for assuming independence via the model-based (naive) variance-covariance matrix of the estimated regression coefficients. The model-based variance is equal to the outside of the robust variance estimator, $\boldsymbol{M}_0^{-1}$, or $\left[\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right]^{-1}$. In this case, the naive variance estimate is computed as if the independent working correlation assumption were correct. In other words, these are the results that would be obtained if clustering were ignored altogether. Although it is not recommended for analysis of clustered data, we are showing it to demonstrate the effects of clustering. We specify ***R=Independent*** (default) and ***SEMETHOD=Model*** on the PROC statement to obtain the independence results.

Exhibit 12. SAS-Callable REGRESS Code for Independence

```
PROC REGRESS DATA=in.boric FILETYPE=SAS R=INDEPENDENT SEMETHOD=MODEL;
  NEST _ONE_ DAMID;
  WEIGHT _ONE_;

  REFLEVEL DOSEGRP = 1;
  CLASS DOSEGRP;
  MODEL BW = DOSEGRP;
  TEST WALDCHI;
  CONDMARG DOSEGRP;

  SETENV COLSPCE=1 LABWIDTH=25 COLWIDTH=8 DECWIDTH=4;
  PRINT BETA="BETA" SEBETA="S.E." T_BETA="T:BETA=0" P_BETA="P-VALUE" /
        TESTS=default t_betafmt=F8.2 dffmt=F7.0 waldchifmt=F10.2 waldchpfmt=f7.4;

  SETENV LABWIDTH=20 decwidth=4 colwidth=6;
  PRINT / COND_MRG=default condmrgfmt=f11.4 t_cndmrgfmt=f8.2 p_cndmrgfmt=f7.4;

  RFORMAT dosegrp dose.;
  RTITLE "Treatment Effect on Fetal Body Weight in a Teratology Experiment";
```

Exhibit 13. First Page of REGRESS Output: Independence

```
                              S U D A A N
           Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute     February 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: DAMID


Number of observations read       :   1302    Weighted count:     1302
Observations used in the analysis :   1302    Weighted count:     1302
Denominator degrees of freedom    :    163

Maximum number of estimable parameters for the model is  6

File IN.BORIC contains  164 Clusters
 164 clusters were used to fit the model
Maximum cluster size is  14 records
Minimum cluster size is   2 records

Weighted mean response is 3.341329

Multiple R-Square for the dependent variable BW: 0.171089
```

Exhibit 14. Linear Regression Coefficients: Independence

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Model-Based (Naive)
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
------------------------------------------------------------------------
Independent Variables and
  Effects                        BETA       S.E.   T:BETA=0    P-VALUE
------------------------------------------------------------------------
Intercept                      3.4979     0.0233     150.15     0.0000
DOSEGRP
  1 = Control                  0.0000     0.0000          .          .
  2 = 0.025%                  -0.0611     0.0332      -1.84     0.0676
  3 = 0.05%                   -0.0789     0.0330      -2.39     0.0180
  4 = 0.075%                  -0.1219     0.0327      -3.73     0.0003
  5 = 0.1%                    -0.2062     0.0323      -6.39     0.0000
  6 = 0.2%                    -0.4883     0.0336     -14.54     0.0000
------------------------------------------------------------------------
```

Exhibit 14 shows that the estimated standard errors assuming independence are much smaller than with the robust variance estimator (***Exhibit 9***), with several of the lower dose groups appearing significantly different from the control. These estimates are overly optimistic (naive) and computed as if the data were truly independent. Therefore, these results are not valid for the data at hand. They merely demonstrate the consequences of ignoring the experimental design.

Exhibit 15. ANOVA Table: Independence

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Model-Based (Naive)
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-------------------------------------------------------------
Contrast                    Degrees                 P-value
                            of                      Wald
                            Freedom   Wald ChiSq    ChiSq
-------------------------------------------------------------
OVERALL MODEL                     6    123708.96     0.0000
MODEL MINUS INTERCEPT             5       267.50     0.0000
INTERCEPT                         .            .          .
DOSEGRP                           5       267.50     0.0000
-------------------------------------------------------------
```

Exhibit 15 contains the main effects tests assuming the naive assumption of independence were true. The *p*-value according to the treatment effect is still significant (*p*=0.0000), but the Wald chi-square is much larger (267.5) than under the robust variance approach.

Exhibit 16. Conditional Marginals: Independence

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Model-Based (Naive)
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-------------------------------------------------------------------------------------
Conditional Marginal
  #1                                          Lower     Upper
                       Conditional            95%       95%
                       Marginal          SE   Limit     Limit     T:Marg=0   P-value
-------------------------------------------------------------------------------------
DOSEGRP
  1 = Control             3.4979   0.0233   3.4519    3.5439      150.15     0.0000
  2 = 0.025%              3.4367   0.0237   3.3900    3.4835      145.13     0.0000
  3 = 0.05%               3.4190   0.0234   3.3728    3.4652      146.09     0.0000
  4 = 0.075%              3.3760   0.0230   3.3306    3.4213      146.91     0.0000
  5 = 0.1%                3.2916   0.0223   3.2475    3.3357      147.36     0.0000
  6 = 0.2%                3.0096   0.0242   2.9618    3.0574      124.34     0.0000
-------------------------------------------------------------------------------------
```

Under independence, the conditional marginals in ***Exhibit 16*** are the same as earlier (see ***Exhibit 11***), but their estimated standard errors and associated confidence limits are misleadingly small and narrow, computed as if the independence assumption were true.

GEE Under Exchangeable Working Correlations and Robust Variance Estimator

Exhibit 17 contains the programming statements for estimating the linear model under exchangeable working correlations. The only change from the previous statements is the switch to ***R=Exchangeable*** on the PROC statement. The default robust variance estimator will be automatically applied (***SEMETHOD=Binder***). All other statements remain unchanged. By default, SUDAAN will use the GEE one-step approach for estimating regression parameters, with the independence parameter estimates being updated exactly once with the estimated correlation structure.

Exhibit 17. SAS-Callable REGRESS Code: GEE Exchangeable

```
PROC REGRESS DATA=in.boric FILETYPE=SAS R=EXCHANGEABLE;
  NEST _ONE_ DAMID;
  WEIGHT _ONE_;

  REFLEVEL DOSEGRP = 1;
  CLASS DOSEGRP;
  MODEL BW = DOSEGRP;
  TEST WALDCHI;
  CONDMARG DOSEGRP;

  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" / TESTS=default RHOS=ALL t_betafmt=F8.2 deftfmt=F6.2
        dffmt=F7.0 waldchifmt=F10.2 waldchpfmt=f7.4;

  SETENV LABWIDTH=20 decwidth=4 colwidth=6;
  PRINT / COND_MRG=default condmrgfmt=f11.4 t_cndmrgfmt=f8.2 p_cndmrgfmt=f7.4;

  RFORMAT dosegrp dose.;
  RTITLE "Treatment Effect on Fetal Body Weight in a Teratology Experiment";
```

Exhibit 18. First Page of REGRESS Output: GEE Exchangeable

```
                                S U D A A N
           Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute     February 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: DAMID
    Cluster Identification Variables: _ONE_ DAMID


Number of observations read       :   1302    Weighted count:     1302
Observations used in the analysis :   1302    Weighted count:     1302
Denominator degrees of freedom    :    163

Maximum number of estimable parameters for the model is  6

File IN.BORIC contains  164 Clusters
 164 clusters were used to fit the model
Maximum cluster size is  14 records
Minimum cluster size is   2 records

Weighted mean response is 3.341329
Multiple R-Square for the dependent variable BW: 0.169118
```

Exhibit 18 indicates that there are 164 clusters, 1,302 fetuses, and a minimum and maximum cluster size of two and 14, respectively.

Exhibit 19 presents the estimated regression coefficients under exchangeability, with a robust variance estimator. We see that these results are qualitatively the same as working independence shown previously in ***Exhibit 7***. Modeling the within-cluster covariance structure has not improved efficiency in these data.

Exhibit 19. Linear Regression Coefficients: GEE Exchangeable

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Exchangeable
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-------------------------------------------------------------------------------
Independent Variables and
  Effects                                        Design
                                  BETA      S.E.  Effect   T:BETA=0    P-VALUE
-------------------------------------------------------------------------------
Intercept                       3.5125    0.0470    0.82      74.67     0.0000
DOSEGRP
  1 = Control                   0.0000    0.0000       .          .          .
  2 = 0.025%                   -0.0509    0.0622    0.75      -0.82     0.4148
  3 = 0.05%                    -0.0656    0.0738    1.02      -0.89     0.3753
  4 = 0.075%                   -0.1363    0.0790    1.19      -1.73     0.0864
  5 = 0.1%                     -0.2409    0.0680    0.89      -3.54     0.0005
  6 = 0.2%                     -0.4822    0.0651    0.78      -7.40     0.0000
-------------------------------------------------------------------------------
```

Exhibit 20, ***Exhibit 21*** and ***Exhibit 22*** contain the main effects tests, correlation parameter, and least squares means under exchangeability with a robust variance estimator. Again, these results are similar to working independence shown earlier.

Exhibit 20. ANOVA Table: GEE Exchangeable

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Exchangeable
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-------------------------------------------------------------
Contrast                    Degrees                 P-value
                            of                      Wald
                            Freedom   Wald ChiSq    ChiSq
-------------------------------------------------------------
OVERALL MODEL                     6     28271.55     0.0000
MODEL MINUS INTERCEPT             5        75.41     0.0000
INTERCEPT                         .           .          .
DOSEGRP                           5        75.41     0.0000
-------------------------------------------------------------
```

Exhibit 21. Estimated Exchangeable Correlation

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Exchangeable
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-----------------------------------
Fetal Body Weight           Correla-
                            tion
                            Matrix
-----------------------------------
1                             0.5056
-----------------------------------
```

The estimated exchangeable correlation parameter (measure of dependence within clusters) is 0.5056 (***Exhibit 21***). The relatively large size of the intracluster correlation is partly responsible for the large design effects (variance inflation) for estimated means and regression parameters seen already. Variance inflation is directly related to the size of the intracluster correlation and the average cluster size (here, number of fetuses per litter).

Exhibit 22. Conditional Marginals: GEE Exchangeable

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Exchangeable
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
--------------------------------------------------------------------------------
Conditional Marginal
  #1                                          Lower    Upper
                      Conditional             95%      95%
                      Marginal          SE    Limit    Limit    T:Marg=0   P-value
--------------------------------------------------------------------------------
DOSEGRP
  1 = Control              3.5125   0.0470   3.4196   3.6054      74.67    0.0000
  2 = 0.025%               3.4616   0.0408   3.3811   3.5421      84.94    0.0000
  3 = 0.05%                3.4469   0.0568   3.3347   3.5592      60.64    0.0000
  4 = 0.075%               3.3762   0.0635   3.2508   3.5015      53.17    0.0000
  5 = 0.1%                 3.2716   0.0491   3.1747   3.3684      66.69    0.0000
  6 = 0.2%                 3.0303   0.0450   2.9414   3.1192      67.31    0.0000
--------------------------------------------------------------------------------
```

GEE Under Exchangeable Working Correlations and a Model-Based (Naive) Variance Estimator

Exhibit 23 contains the programming statements for the exchangeable correlation model using the model-based (naive) variance-covariance matrix of the estimated regression coefficients. The model-based variance is the $\boldsymbol{M}_0^{-1}$ matrix, or the outside portion of the robust variance estimate: $\boldsymbol{M}_0^{-1} = \left[\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right]^{-1}$. In this case, the naive variance estimate is computed assuming that the exchangeable working correlation assumptions were correct. Since that is close to truth for litter data, we will see that results are essentially the same as with the robust variance estimator. In the SUDAAN code, we specify ***R=Exchangeable*** and ***SEMETHOD=Model***.

Exhibit 23. SAS-Callable REGRESS Code: GEE Exchangeable with Model-Based Variance

```
PROC REGRESS DATA=in.boric FILETYPE=SAS R=EXCHANGEABLE SEMETHOD=MODEL;
  NEST _ONE_ DAMID;
  WEIGHT _ONE_;

  REFLEVEL DOSEGRP = 1;
  CLASS DOSEGRP;
  MODEL BW = DOSEGRP;
  TEST WALDCHI;
  CONDMARG DOSEGRP;

  SETENV COLSPCE=1 LABWIDTH=25 COLWIDTH=8 DECWIDTH=4;
  PRINT BETA="BETA" SEBETA="S.E." T_BETA="T:BETA=0" P_BETA="P-VALUE" /
        TESTS=default t_betafmt=F8.2 dffmt=F7.0 waldchifmt=F10.2 waldchpfmt=f7.4;

  SETENV LABWIDTH=20 decwidth=4 colwidth=6;
  PRINT / COND_MRG=default condmrgfmt=f11.4 t_cndmrgfmt=f8.2 p_cndmrgfmt=f7.4;

  RFORMAT dosegrp dose.;
  RTITLE  "Treatment Effect on Fetal Body Weight in a Teratology Experiment";
```

Exhibit 24. First Page of REGRESS Output: GEE Exchangeable with Model-Based Variance

```
                                S U D A A N
           Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute     February 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: DAMID
    Cluster Identification Variables: _ONE_ DAMID

Number of observations read       :   1302    Weighted count:     1302
Observations used in the analysis :   1302    Weighted count:     1302
Denominator degrees of freedom    :    163

Maximum number of estimable parameters for the model is  6

File IN.BORIC contains  164 Clusters
 164 clusters were used to fit the model
Maximum cluster size is  14 records
Minimum cluster size is   2 records

Weighted mean response is 3.341329

Multiple R-Square for the dependent variable BW: 0.169118
```

Exhibit 25. Linear Regression Coefficients: GEE Exchangeable with Model-Based Variance

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Model-Based (Naive)
Working Correlations: Exchangeable
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-------------------------------------------------------------------------
Independent Variables and
  Effects                        BETA       S.E.    T:BETA=0    P-VALUE
-------------------------------------------------------------------------
Intercept                      3.5125     0.0519      67.69     0.0000
DOSEGRP
  1 = Control                  0.0000     0.0000         .          .
  2 = 0.025%                  -0.0509     0.0719      -0.71     0.4799
  3 = 0.05%                   -0.0656     0.0730      -0.90     0.3701
  4 = 0.075%                  -0.1363     0.0723      -1.88     0.0613
  5 = 0.1%                    -0.2409     0.0721      -3.34     0.0010
  6 = 0.2%                    -0.4822     0.0737      -6.54     0.0000
-------------------------------------------------------------------------
```

Exhibit 25 contains the estimated regression coefficients computed under exchangeability and the standard errors as if the exchangeable working assumption were correct. Exchangeability is a reasonable assumption for data on littermates. The standard errors are roughly the same as with the robust variance estimator for these data, indicating that the exchangeable correlation assumption is close to truth.

Exhibit 26. ANOVA Table: GEE Exchangeable with Model-Based Variance

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Model-Based (Naive)
Working Correlations: Exchangeable
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-----------------------------------------------------------
Contrast                   Degrees                 P-value
                           of                      Wald
                           Freedom   Wald ChiSq    ChiSq
-----------------------------------------------------------
OVERALL MODEL                    6     26050.93     0.0000
MODEL MINUS INTERCEPT            5        58.46     0.0000
INTERCEPT                        .           .          .
DOSEGRP                          5        58.46     0.0000
-----------------------------------------------------------
```

Exhibit 26 contains the main effects tests computed under exchangeability, using the model-based variance approach. Results are essentially the same as with the robust variance estimator (***Exhibit 20***).

Exhibit 27. Conditional Marginals: GEE Exchangeable with Model-Based Variance

```
Variance Estimation Method: Taylor Series (WR)
SE Method: Model-Based (Naive)
Working Correlations: Exchangeable
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
--------------------------------------------------------------------------------
Conditional Marginal
  #1                                        Lower    Upper
                      Conditional           95%      95%
                      Marginal          SE  Limit    Limit    T:Marg=0   P-value
--------------------------------------------------------------------------------
DOSEGRP
  1 = Control              3.5125   0.0519  3.4100   3.6150      67.69    0.0000
  2 = 0.025%               3.4616   0.0497  3.3635   3.5598      69.64    0.0000
  3 = 0.05%                3.4469   0.0513  3.3456   3.5482      67.17    0.0000
  4 = 0.075%               3.3762   0.0504  3.2766   3.4757      66.97    0.0000
  5 = 0.1%                 3.2716   0.0501  3.1727   3.3705      65.33    0.0000
  6 = 0.2%                 3.0303   0.0523  2.9270   3.1336      57.92    0.0000
--------------------------------------------------------------------------------
```

Exhibit 27 contains the conditional marginals computed under exchangeability, using the model-based variance approach. Marginals are identical (as expected) and SEs and confidence limits are comparable to those computed with the robust variance estimator (***Exhibit 22***).

Jackknife Variance Estimation

Exhibit 28 contains the modeling code for an alternative approach, Delete-1 Jackknife variance estimation (a replication technique). We obtained these results by specifying ***DESIGN=Jackknife*** on the PROC statement.

Exhibit 28. SAS-Callable REGRESS Code: Jackknife Variance

```
PROC REGRESS DATA=in.boric FILETYPE=SAS R=INDEPENDENT DESIGN=JACKKNIFE;
  NEST _ONE_ DAMID;
  WEIGHT _ONE_;

  REFLEVEL DOSEGRP = 1;
  CLASS DOSEGRP;
  MODEL BW = DOSEGRP;
  CONDMARG DOSEGRP;
  TEST WALDCHI;

  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" / TESTS=default t_betafmt=F8.2 deftfmt=F6.2 dffmt=F7.0
        waldchifmt=F10.2 waldchpfmt=f7.4;

  SETENV LABWIDTH=20 decwidth=4 colwidth=6;
  PRINT / COND_MRG=default condmrgfmt=f11.4 t_cndmrgfmt=f8.2 p_cndmrgfmt=f7.4;

  RFORMAT dosegrp dose.;
  RTITLE "Treatment Effect on Fetal Body Weight in a Teratology Experiment";
```

Exhibit 29. First Page of REGRESS Output: Jackknife Variance

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

DESIGN SUMMARY: Variances will be computed using the Delete-1 Jackknife (JACKKNIFE)
Replication Method
    Sample Weight: _ONE_
    Stratification Variable(s): _ONE_
    Primary Sampling Unit (PSU): DAMID

Number of observations read       :   1302    Weighted count:     1302
Observations used in the analysis :   1302    Weighted count:     1302
Denominator degrees of freedom    :    163

Maximum number of estimable parameters for the model is  6

File IN.BORIC contains  164 Clusters
 164 clusters were used to fit the model
Maximum cluster size is  14 records
Minimum cluster size is   2 records

Weighted mean response is 3.341329
Multiple R-Square for the dependent variable BW: 0.171089
```

Exhibit 30. Linear Regression Coefficients: Jackknife Variance

```
Variance Estimation Method: Delete-1 Jackknife
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
------------------------------------------------------------------------------
Independent Variables and
  Effects                                        Design
                                BETA       S.E.  Effect   T:BETA=0    P-VALUE
------------------------------------------------------------------------------
Intercept                     3.4979     0.0419    3.23      83.52     0.0000
DOSEGRP
  1 = Control                 0.0000     0.0000     .          .         .
  2 = 0.025%                 -0.0611     0.0616    3.44      -0.99     0.3227
  3 = 0.05%                  -0.0789     0.0752    5.18      -1.05     0.2955
  4 = 0.075%                 -0.1219     0.0765    5.46      -1.59     0.1128
  5 = 0.1%                   -0.2062     0.0648    4.03      -3.18     0.0018
  6 = 0.2%                   -0.4883     0.0624    3.45      -7.83     0.0000
------------------------------------------------------------------------------
```

Exhibit 30 contains the estimated regression coefficients and standard errors under the jackknife variance estimator. Note that the regression coefficient estimates are simply those computed under independence. They are unaffected by the specification of the jackknife option. The estimated standard errors are computed using the jackknife variance estimator. Note that the results for this example are similar to those obtained under GEE-independent and GEE-exchangeable.

Exhibit 31 contains the main effects tests using the jackknife approach. Again, they are similar to those obtained under GEE.

Exhibit 31. ANOVA Table: Jackknife Variance

```
Variance Estimation Method: Delete-1 Jackknife
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
------------------------------------------------------------
Contrast                    Degrees                P-value
                            of                     Wald
                            Freedom   Wald ChiSq   ChiSq
------------------------------------------------------------
OVERALL MODEL                     6     27208.92    0.0000
MODEL MINUS INTERCEPT             5        72.72    0.0000
INTERCEPT                         .            .         .
DOSEGRP                           5        72.72    0.0000
------------------------------------------------------------
```

Exhibit 32 contains the conditional marginals (least squares means) and their standard errors using the Delete-1 jackknife approach. Marginals are identical to those computed under independence, and the SEs and confidence limits are comparable to those computed under the other robust variance estimators.

Exhibit 32. Conditional Marginals: Jackknife Variance

```
Variance Estimation Method: Delete-1 Jackknife
Working Correlations: Independent
Link Function: Identity
Response variable BW: Fetal Body Weight

Treatment Effect on Fetal Body Weight in a Teratology Experiment
-------------------------------------------------------------------------------
Conditional Marginal
  #1                                        Lower    Upper
                      Conditional             95%      95%
                      Marginal          SE    Limit    Limit    T:Marg=0   P-value
-------------------------------------------------------------------------------
DOSEGRP
  1 = Control              3.4979   0.0419   3.4152   3.5806      83.52    0.0000
  2 = 0.025%               3.4367   0.0452   3.3475   3.5260      76.05    0.0000
  3 = 0.05%                3.4190   0.0624   3.2957   3.5422      54.77    0.0000
  4 = 0.075%               3.3760   0.0640   3.2496   3.5023      52.77    0.0000
  5 = 0.1%                 3.2916   0.0495   3.1939   3.3893      66.54    0.0000
  6 = 0.2%                 3.0096   0.0462   2.9183   3.1009      65.09    0.0000
-------------------------------------------------------------------------------
```