Chapter 5 A Repeated Measures ANOVA with Temporally-Correlated Errors

Purpose:

- Illustrate how **correlated** measurements are modeled.
- Illustrate how the PROC MIXED **REPEATED** statement is used to **simultaneously model** a covariance structure consisting of both an **Independent** Subject/Block Effect **and Correlated** Repeated Measurements within Subjects/Blocks.
- Illustrate Steps in Choosing an Appropriate Covariance Structure

Introduction

The Split-Plot ANOVA in Chapter 4 illustrated how increased precision can be obtained by subdividing experimental units "whole-units" into "sub-units or split-units" and randomly applying the levels of the treatment factor of greatest import/interest to the sub-units. Sometimes, however, random assignment to sub-units is not possible due to temporal and/or spatial constraints. In other words, the treatment factor to be applied to sub-units is "Time of Measurement" or "Location". **"Time" and "Location" can obviously NOT be randomly assigned** because they naturally occur when or where they do. **This fact makes these sub-units potentially "dependent" upon** (i.e., correlated with) **one another**. The following example illustrates how to fit a mixed model when at least one treatment factor exhibits correlation among its levels.

The steps in the following process are outlined in Appendix F.

AN EXAMPLE – REPEATED MEASURES IN A CRD SPLIT-PLOT

Also known as a **split-plot in time**. The whole-plot factor is arranged as a CRD. The whole plots are subjects. The subplot factor is time. **Time is arranged within a block design** (not randomized and not necessarily complete) where the **blocks are subjects**. The **whole-plot factor and time are** considered **fixed** and the **subjects are** considered **random**.

Step 1. Research Objectives:

A condition in a particular breed of dairy cow has been identified as a cause of reduction in milk production potential. It is hypothesized that the supplementation of diet with one or both of two experimental treatments, IA and SC each at levels of 0 or 20 units will effect a significant reduction in the biological marker for this medical condition. This reduction will occur within a 21-day time period, after which it has been determined to be irreversible.

Step 2. Identify and Define Experiment Design:

2.a. Experimental Units

- Whole-unit is a dairy cow
- Sub-unit is a sample taken from a cow at a specific time

2.b. Treatments:

2 x 2 Factorial: (IA= 0, SC=0) (IA= 0, SC=20)
(IA=20, SC=0) (IA=20, SC=20)

2.c. Treatment Application to Experimental Units:

Physiological function indicates that changes in the biological indicator require at least 4 days and that the majority of change will probably occur within the first 10 days. Resource constraints for taking and processing samples limited the study to 4 replicate cows per treatment combination and 4 repeated measurements sampled from each cow at 3, 6, 9, and 21-days post diet introduction.

The 16 cows are randomly divided into 4 groups of 4 and each of the 4 (2x2 factorial) treatment combinations of IA and SC are assigned to one group of cows. Cows are whole-units (see below Experiment Layout diagram). Since **each cow (i.e., whole-unit) is assigned a treatment in a completely randomized manner** and each cow measured at 4 pre-specified amounts of time from the experiment's beginning, the design can be called: Completely Randomized with Repeated Measurements in Time. Note, this is **a similar design to** the split-unit design in **Chapter 4**. The only difference is that the **time period** (i.e., sub-unit) in a repeated-measures design **cannot be randomized** as the sub-unit can be in a split-unit design.

2.d. Response Measurements Y = Biological Marker

Experiment Layout:

		(IA=0, SC=20)				(IA=20, SC=20)				(IA=20, SC=0)				(IA=0, SC=0)			
Whole-Plot Factor(**Cow**):		**1**	**8**	**10**	**15**	**2**	**7**	**9**	**16**	**3**	**5**	**12**	**14**	**4**	**6**	**11**	**13**
Subplot Factor (**Days**)	**3**	258	192	234	256	233	152	186	221	260	228	224	197	269	196	202	212
	6	251	185	233	249	236	148	186	219	255	221	217	196	258	191	202	212
	9	245	183	228	237	232	144	185	225	245	221	209	190	249	181	202	212
	21	242	181	219	247	219	139	167	201	249	214	201	281	253	195	189	219

Step 3. Compile Data

Program 5.1:

```
DATA iasc;
 INPUT cow ia$ sc$ day y @@;
  ia_sc= LEFT(RIGHT(ia)||"_"||LEFT(sc));
 trtcombo=LEFT(RIGHT(ia_sc)||LEFT(day));
 DATALINES;
   4  0   0   3  269    4  0   0   6  258    4  0   0   9  249    4  0   0  21  253
   6  0   0   3  196    6  0   0   6  191    6  0   0   9  181    6  0   0  21  195
  11  0   0   3  202   11  0   0   6  202   11  0   0   9  202   11  0   0  21  189
  13  0   0   3  212   13  0   0   6  212   13  0   0   9  212   13  0   0  21  219

   1  0  20   3  258    1  0  20   6  251    1  0  20   9  245    1  0  20  21  242
   8  0  20   3  192    8  0  20   6  185    8  0  20   9  183    8  0  20  21  181
  10  0  20   3  234   10  0  20   6  233   10  0  20   9  228   10  0  20  21  219
  15  0  20   3  256   15  0  20   6  249   15  0  20   9  237   15  0  20  21  247

   3 20   0   3  260    3 20   0   6  255    3 20   0   9  245    3 20   0  21  249
   5 20   0   3  228    5 20   0   6  221    5 20   0   9  221    5 20   0  21  214
```

```
  12 20   0   3  224   12 20   0   6  217   12 20   0   9  209   12 20   0  21  201
  14 20   0   3  197   14 20   0   6  196   14 20   0   9  190   14 20   0  21  281

   2 20  20   3  233    2 20  20   6  236    2 20  20   9  232    2 20  20  21  219
   7 20  20   3  152    7 20  20   6  148    7 20  20   9  144    7 20  20  21  139
   9 20  20   3  186    9 20  20   6  186    9 20  20   9  185    9 20  20  21  167
  16 20  20   3  221   16 20  20   6  219   16 20  20   9  225   16 20  20  21  201
;
RUN;
```

Step 4. Verify Data

Program 5. 2:

```
PROC SUMMARY DATA= a ;
   CLASS  ia sc day ;
   VAR   y ;
   OUTPUT OUT= look N= n   MEAN= mean ;
RUN;
PROC SORT DATA= look ; BY _TYPE_;
PROC PRINT DATA= look   NOOBS;   BY _TYPE_; RUN;
```

Step 5. Specify the Statistical Model

5.a. Identify Fixed & Random Effects:

Fixed Effects: IA SC Day

Random Effects: Cow

5.b. Write the Statistical Model:

WP Error = WP Rep x WP Trt

$$y_{ijkd} = \mu + \alpha_i + \tau_j + \alpha\tau_{ij} + \epsilon_{\mathbf{k(ij)}} \quad \leftarrow \text{cow terms}$$

$$+ \delta_{d(ijk)} + \delta\alpha_{d(ijk)} + \delta\tau_{d(ijk)} + \delta\alpha\tau_{d(ijk)} + \epsilon_{\mathbf{d(ijk)}} \quad \leftarrow \text{day(cow) terms}$$

Residual Error

where

y_{ijkd} is the observed value of the bio-marker on day d for cow k assigned IA treatment i and SC treatment j.

μ is the average of all observed bio-marker values

$\alpha_i + \tau_j + \alpha\tau_{ij}$ are the fixed effects, respectively, of: the IA treatment i, the SC treatment j, and the interaction between the i^{th} IA treatment and the j^{th} SC treatment.

$\epsilon_{\mathbf{k(ij)}}$ **is the random effect of cow k**; assigned IA treatment i and SC treatment j. This is the "whole-unit" error term. $\epsilon_{\mathbf{k(ij)}} \sim$ **i.i.d.** η **(0,** $\sigma^2_{\mathbf{cow}}$**)**

$\delta_{d(ijk)} + \delta\alpha_{d(ijk)} + \delta\tau_{d(ijk)} + \delta\alpha\tau_{d(ijk)}$ are the fixed effects involving the d^{th} day for cow k; assigned IA treatment i and SC treatment j.

$\epsilon_{\mathbf{d(ijk)}}$ **is the random effect of the d^{th} day for cow k**. This is the "sub-unit" or **"residual"** error term. $\epsilon_{\mathbf{d(ijk)}} \sim \eta$ **(0,** $\sigma^2_{\mathbf{day(cow)}}$**)**. The covariance structure, $\sigma^2_{day(cow)}$, of $\epsilon_{d(ijk)}$ is not expressed in more detail because, given the anticipated dependency (i.e., correlation) among observed measurements at subsequent times for the same cow, several structures (Appendix E.2) will be fit to the data to determine the most appropriate model (Appendix E.3).

Therefore, ***Among Cows*** ***Within Cows***

Use MODEL statement to estimate Fixed Effect Means

$$E(y_{ijd}) = \mu + \alpha_i + \tau_j + \alpha\tau_{ij} + \delta_d + \delta\alpha_{d(i)} + \delta\tau_{d(j)} + \delta\alpha\tau_{d(ij)}$$

Eq. 5.1

Use RANDOM &/or REPEATE D statement to estimate Random Effect Variances

$$Var(y_{ijkd}) = \sigma^2_{cow} + \sigma^2_{day(cow)}$$

Eq. 5.2

$$Cov(y_{ijkd}, y_{ijkm}) = \sigma^2_{cow} + cov(\epsilon_{d(ijk)}, \epsilon_{m(ijk)})$$

Eq. 5.3

for Days = m & d; m≠d. Days are within Cow.

For a particular cow, the covariance Structure (i.e., relationship among "repeated measurements" observed on Days= 3, 6, 9, and 21) identified in Equations 5.2 and 5.3 can be written in the following matrix notation, Equation 5.4.

$$R_k = \begin{bmatrix} \sigma^2_{cow} + \sigma^2_{day3(cow)} & \sigma^2_{cow} + \sigma_{(day3,day6)(cow)} & \sigma^2_{cow} + \sigma_{(day3,day9)(cow)} & \sigma^2_{cow} + \sigma_{(day3,day21)(cow)} \\ \sigma^2_{cow} + \sigma_{(day6,day3)(cow)} & \sigma^2_{cow} + \sigma^2_{day6(cow)} & \sigma^2_{cow} + \sigma_{(day6,day9)(cow)} & \sigma^2_{cow} + \sigma_{(day6,day21)(cow)} \\ \sigma^2_{cow} + \sigma_{(day9,day3)(cow)} & \sigma^2_{cow} + \sigma_{(day9,day6)(cow)} & \sigma^2_{cow} + \sigma^2_{day9(cow)} & \sigma^2_{cow} + \sigma_{(day9,day21)(cow)} \\ \sigma^2_{cow} + \sigma_{(day21,day3)(cow)} & \sigma^2_{cow} + \sigma_{(day21,day6)(cow)} & \sigma^2_{cow} + \sigma_{(day21,day9)(cow)} & \sigma^2_{cow} + \sigma^2_{day21(cow)} \end{bmatrix}$$

Eq. 5.4

Rows 1 to 4 and Columns 1 to 4 of this matrix, $\boldsymbol{R}_k$, are associated with Day 3, Day 6, Day 9, and Day 21 measurements, respectively. Variances (Eq. 5.2) are placed on the diagonal of $\boldsymbol{R}_k$ and covariances (Eq. 5.3) are off-diagonal elements of $\boldsymbol{R}_k$. The $\sigma^2_{day\ i\ (cow)}$ variance components in each diagonal element are frequently similar regardless of Day, but if the $\sigma^2_{day\ i\ (cow)}$ are heterogeneous for some Day(s), the GROUP= option of the REPEATED statement can be used to model these distinct covariance parameters (as shown in Chapter 1 - Example 3).

$\boldsymbol{R}_k$ is expressed in Equation 5.4 using notation general enough to represent any one of a wide variety of possible covariance structures (Appendix E.2). Most appropriate covariance structure for this specific data set will be determined in Step 6, below.

5.c. (Optional) - Construct an ANOVA Table using PROC GLM

The purpose of this program is to use PROC GLM to get a general idea of which Mean Squares (or functions thereof) are appropriate denominators to obtain correct F-tests for each fixed effect in the model. Recall that PROC GLM does not generally produce correct analyses for models containing random effects. Hence, the purpose of using PROC GLM here is only to provide a preliminary approximation for the F-tests for each fixed effect that PROC MIXED will provide correctly.

Cow is a Whole-Plot Random Block

Program 5.3:

```
PROC GLM   DATA= iasc ;
  CLASS   ia  sc day cow  ;
  MODEL  y  =  ia|sc|day  cow(ia sc) / SS3 ;
  RANDOM   cow(ia sc)  / TEST ;
RUN ;
```

Output Listing from Program 5.3:

```
                         The GLM Procedure
                      Class Level Information
 Class         Levels    Values
 ia                 2    0 20
 sc                 2    0 20
 day                4    3 6 9 21
 cow               16    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

           Source      Type III Expected Mean Square

           ia          Var(Error) + 4 Var(cow(ia*sc)) + Q(ia,ia*sc,ia*day,ia*sc*day)
           sc          Var(Error) + 4 Var(cow(ia*sc)) + Q(sc,ia*sc,sc*day,ia*sc*day)
           ia*sc       Var(Error) + 4 Var(cow(ia*sc)) + Q(ia*sc,ia*sc*day)
WP Error:  cow(ia*sc)  Var(Error) + 4 Var(cow(ia*sc))

           day         Var(Error) + Q(day,ia*day,sc*day,ia*sc*day)
           ia*day      Var(Error) + Q(ia*day,ia*sc*day)
           sc*day      Var(Error) + Q(sc*day,ia*sc*day)
           ia*sc*day   Var(Error) + Q(ia*sc*day)
SP Error:  Residual    Var(Error)
```

The following "Tests of Hypotheses" are constructed by PROC GLM based on the information it calculates in the above "Expected Mean Squares" listing. PROC MIXED with the DDFM=KR option in the MODEL statement provides the most accurate F-test result. **PROC GLM provides only an understanding of approximately how the F-tests are conducted via the TEST option in the RANDOM statement**, but the **subsequent F-tests provided by PROC GLM are not necessarily correct**. When the data being analyzed is unbalanced relative to the number of replicates per treatment, appropriate denominators for some F-tests are linear functions of two or more random effects. Denominator (i.e., Error) degrees of freedom are also functions of the mean squares and their associated degrees of freedom. When appropriate, the DDFM=KR option in the MODEL statement in PROC MIXED automatically accomplishes these calculations. Details are available under the MODEL statement options in the PROC MIXEDsection of the SAS9 Help and Documentation (link provided in Appendix C.7).

```
                              The GLM Procedure
          Tests of Hypotheses for Mixed Model Analysis of Variance


     Source                    DF     Type III SS     Mean Square    F Value    Pr > F
*    ia                         1     2268.140625     2268.140625       0.64    0.4378
*    sc                         1     1570.140625     1570.140625       0.45    0.5169
*    ia*sc                      1     7943.265625     7943.265625       2.26    0.1590
     Error: MS(cow(ia*sc))     12           42252     3521.036458
* This test assumes one or more other fixed effects are zero.


     Source                    DF     Type III SS     Mean Square    F Value    Pr > F
*    day                        3      619.296875      206.432292       1.25    0.3052
*    ia*day                     3       69.671875       23.223958       0.14    0.9348
*    sc*day                     3      970.921875      323.640625       1.96    0.1369
     ia*sc*day                  3      531.546875      177.182292       1.08    0.3719
     cow(ia*sc)                12           42252     3521.036458      21.36    <.0001
     Error: MS(Error)          36     5933.312500      164.814236
* This test assumes one or more other fixed effects are zero.
```

Step 5.d. Write PROC MIXED Code

As mentioned above, for this Repeated Measures ANOVA the whole-units (i.e., **cows) are considered to be independent of one another**, but the sub-units (i.e., **time periods) will very likely exhibit a dependence structure**. Hence, the PROC MIXED REPEATED statement will be used, in the process outlined in Appendix E.3, to identify and estimate the covariance structure present in the observed data.

Program 5.4a **- Determine the Appropriate Covariance Structure**

Goal: Find covariance structure

```
PROC MIXED DATA = iasc ;
  CLASS cow ia sc day ;
  MODEL y = ia sc ia*sc
            day day*ia day*sc day*ia*sc / DDFM= KR OUTPRED=resids;
  REPEATED /SUBJECT=cow(ia sc) TYPE=< > R RCORR;
TITLE 'Covariance Structure is: Name of the Covariance Structure';
QUIT;
```

Note absence of a RANDOM statement in program 5.4a. There is no “block” for cows. The SUBJECT= option in the REPEATED statement is used to identify the “whole” experimental units (i.e., cows). All “repeated” measurements recorded on a cow will likely exhibit dependency. The TYPE= option allows choice among several covariance structures to model this dependency.

Step 6. Fine-Tune the Model Specification

6. a. Is the Data Correlated in Time or Space ?

Obviously, since 4 measurements were made as a **time** series on each of the cows, **there is potential for dependence (i.e., correlation) among** these measurements.

6. a. i. Define & Fit Candidate Correlation Structures

The following procedure is recommended for choosing the most appropriate covariance structure for your data:

1) **Fit TYPE=UN first** (if PROC MIXED is able to fit this very complex structure to your specific data set).

2) **View the Covariance Parameter Matrix** resulting from the **TYPE=UN** fit and/or **determine** philosophically which of the **candidate covariance structures** would reasonably describe the actual correlation present in the data. For a listing of covariance structures available in SAS PROC MIXED, **see Appendix E.2 or the SAS doc link in Appendix C.9.**

3) Fit the candidate covariance structures using Programs 5.4a & 5.4b. Compare the values of the "Fit Statistics" among the candidate covariance structures. The covariance structure yielding the **smallest values for the "Fit Statistics"** (eg., AICC) provides the **most accurate fit to the correlation structure** that exists in your data.

   Note: The "Null Model Likelihood Ratio Test" (i.e., LRT) does not allow comparison among all candidate covariance structures. The LRT tests only whether the candidate covariance is a significantly better fit to the data than TYPE=VC (i.e., assuming no dependency).

4) Run Program 5.5 using the covariance structure selected in 3) and specify all significant fixed effects in the LSMEANS statement to obtain means comparisons.

>>> Since the **potential dependence is among 4 measurements over time on the same cow**, it makes sense to **consider covariance structures** that are **'regressive' in nature**. Two such structures are **ANTE(1)** and **SP(EXP)**. We will also fit the VC (i.e., independence - no correlation - only variance components) structure to confirm that there is significant correlation among time measurements. The **estimated covariance matrix from the TYPE=UN analysis**, shown below, **suggests that a CS or CSH** covariance structure **may** also **be reasonable** since the off-diagonal covariances do not differ markedly from one another (CS) and the diagonal covariances seem to decrease in size as going down the diagonal (CSH).

Step 6. Fine-Tune the Model Specification

6. a. i. Define & Fit Candidate Correlation Structures (cont'd)

What follows is the "Fit Statistics" portion of Program 5.4a for each of these candidate covariance structures. Small values of the Fit Statistics indicate the better model fit to the data. For illustrative purposes, since this is the first discussion of fitting several different covariance structures, "Tests for Fixed Effects" are listed for each covariance structure, to emphasize the effect that different models of the random effects have upon fixed effect inferences. Typically, Program 5.4b can be used to facilitate identification of the appropriate covariance structure and minimize the amount of SAS output.

Output Listing for Program 5.4a - TYPE=UN:

```
               CRD Repeated Measures
              Covariance Type is  UN

                      Dimensions
        Covariance Parameters        10

      Estimated R Matrix for cow(ia*sc) 4 0 0
Row        Col1        Col2        Col3        Col4    Note: TYPE=UN
  1     1016.46      991.29      955.40      714.98     allows each of the
  2      991.29      979.44      948.73      730.31     4 variances and
  3      955.40      948.73      940.21      693.63     6 covariances to be
  4      714.98      730.31      693.63     1079.38     estimated individually.

          Null Model Likelihood Ratio Test
            DF    Chi-Square      Pr > ChiSq
             9        108.87          <.0001

                  Fit Statistics
       -2 Res Log Likelihood          381.3
       AIC (smaller is better)        401.3
       AICC (smaller is better)       407.2
       BIC (smaller is better)        409.0

          Type 3 Tests of Fixed Effects
                Num     Den
  Effect         DF      DF    F Value    Pr > F
  ia              1      12       0.64    0.4378
  sc              1      12       0.45    0.5169
  ia*sc           1      12       2.26    0.1590
  day             3      10       6.83    0.0087
  ia*day          3      10       0.40    0.7545
  sc*day          3      10       1.04    0.4159
  ia*sc*day       3      10       1.52    0.2696
```

The highly significant "Null Likelihood Ratio Test" (LRT) indicates presence of significant dependency (i.e., correlation). Below analyses fit covariance

Step 6. Fine-Tune the Model Specification

6. a. i. Define & Fit Candidate Correlation Structures (cont'd)

structures that are simpler (i.e., require fewer unique covariance parameters to be estimated). The objective is to fit as simple a covariance structure as possible that accurately models the data's dependency structure. This retains

the maximum degrees of freedom for the error term, which provides greatest statistical power of the F-tests.

Output Listing for Program 5.4a - TYPE=VC:

```
              Covariance Type is  VC

                       Dimensions
        Covariance Parameters         1

             Convergence criteria met.

      Estimated R Matrix for cow(ia*sc) 4 0 0
Row        Col1        Col2        Col3        Col4
  1     1003.87                                          Note: TYPE=VC
  2                 1003.87                                  requires estimate
  3                             1003.87                      of a single
  4                                         1003.87          variance

                  Fit Statistics
       -2 Res Log Likelihood          490.2
       AIC (smaller is better)        492.2
       AICC (smaller is better)       492.2
       BIC (smaller is better)        492.9


         Null Model Likelihood Ratio Test
           DF    Chi-Square      Pr > ChiSq
            0          0.00          1.0000

          Type 3 Tests of Fixed Effects
                  Num     Den
  Effect          DF      DF     F Value    Pr > F
  ia               1      48        2.26    0.1394
  sc               1      48        1.56    0.2171
  ia*sc            1      48        7.91    0.0071
  day              3      48        0.21    0.8920
  ia*day           3      48        0.02    0.9952
  sc*day           3      48        0.32    0.8091
  ia*sc*day        3      48        0.18    0.9118
```

TYPE=VC assumes there is no covariance. Obviously, the covariance structure estimated above by TYPE=UN indicates there is non-zero covariance among the 4

Step 6. Fine-Tune the Model Specification

6. a. i. Define & Fit Candidate Correlation Structures (cont'd)

time measurements. Hence, it is no surprise that the "Fit Statistics" are much larger (indicating an inadequate model fit to the data) for TYPE=VC than for TYPE=UN.

Of all covariance structures, TYPE=UN estimates the largest number of covariances and TYPE=VC estimates the least (just one).

The next logical step is to assume that all covariances are equal and fit the structure TYPE=CS.

Output Listing for Program 5.4a - TYPE=CS:

```
                     Dimensions
       Covariance Parameters         2

              Covariance Type is  CS

     Estimated R Matrix for cow(ia*sc) 4 0 0
Row        Col1         Col2         Col3         Col4
  1     1003.87       839.06       839.06       839.06   Note: TYPE=CS
  2      839.06      1003.87       839.06       839.06    requires estimates
  3      839.06       839.06      1003.87       839.06    for 1 variance
  4      839.06       839.06       839.06      1003.87    and 1 covariance.

                 Fit Statistics
      -2 Res Log Likelihood          440.2         Note: Type=CS
      AIC (smaller is better)        444.2         Fit Statistics values
      AICC (smaller is better)       444.4         fall between those of
      BIC (smaller is better)        445.7         Type=UN and Type=VC

        Null Model Likelihood Ratio Test
          DF    Chi-Square      Pr > ChiSq
           1         49.99          <.0001

         Type 3 Tests of Fixed Effects
              Num     Den
  Effect        DF      DF    F Value    Pr > F
  ia             1      12       0.64    0.4378
  sc             1      12       0.45    0.5169
  ia*sc          1      12       2.26    0.1590
  day            3      36       1.25    0.3052
  ia*day         3      36       0.14    0.9348
  sc*day         3      36       1.96    0.1369
  ia*sc*day      3      36       1.08    0.3719
```

TYPE=CS estimates the covariances to be 839.06. The "Fit Statistics" are much smaller than for TYPE=VC, indicating TYPE=CS fits the data substantially better. Significance of the "Null Model Likelihood Ratio Test" is also an indicator TYPE=CS model fits the data better than TYPE=VC.

Step 6. Fine-Tune the Model Specification

6. a. i. Define & Fit Candidate Correlation Structures (cont'd)

Since the variances estimates in the TYPE=UN case (i.e., covariances on the diagonal of the R matrix) appear to be decreasing in size, it would be instructive to try and fit a structure that allows this relationship: TYPE=CSH.

Output Listing for Program 5.4a - TYPE=CSH:

```
                          Dimensions
              Covariance Parameters          5

                     Covariance Type is  CSH

            Estimated R Matrix for cow(ia*sc) 4 0 0
Row          Col1         Col2         Col3         Col4
  1        960.11       791.85       781.56       958.15   Note: TYPE=CSH
  2        791.85       911.52       761.52       933.58    requires estimates
  3        781.56       761.52       887.98       921.45    of 5 covariances (see
  4       958.15       933.58       921.45      1334.58     Table 46.5 in SAS Doc.
                                                            via link in App C.7)

                        Fit Statistics
             -2 Res Log Likelihood           438.1
             AIC (smaller is better)         448.1
             AICC (smaller is better)        449.6
             BIC (smaller is better)         452.0

               Null Model Likelihood Ratio Test
                 DF    Chi-Square      Pr > ChiSq
                  4         52.03          <.0001

                Type 3 Tests of Fixed Effects
                      Num     Den
  Effect              DF      DF     F Value    Pr > F
  ia                   1      12        0.64    0.4409
  sc                   1      12        0.44    0.5198
  ia*sc                1      12        2.23    0.1616
  day                  3    26.7        1.28    0.3005
  ia*day               3    26.7        0.12    0.9449
  sc*day               3    26.7        1.42    0.2577
  ia*sc*day            3    26.7        0.82    0.4930
```

The larger values for the "Fit Statistics" indicate that TYPE=CSH is not as good a fit to the data as was TYPE=CS. The significant "Null Model Likelihood Ratio Test" continues to indicate that the TYPE=CSH model provides a statistically better fit to the data than does the TYPE=VC model (which assumes no dependency structure). The "Fit Statistics" indicate the **TYPE=CS model provides the best fit to the data, thusfar**.

Step 6. Fine-Tune the Model Specification

6. a. i. Define & Fit Candidate Correlation Structures (cont'd)

Since the correlation is between time measurements, let's next try covariance structures developed specifically for temporal (i.e., ANTE(1)) and spatial (i.e., SP(EXP)) relationships.

Output Listing for Program 5.4a - TYPE=ANTE(1):

```
                    Dimensions
       Covariance Parameters          7

            Covariance Type is  ANTE(1)

     Estimated R Matrix for cow(ia*sc) 4 0 0
Row        Col1         Col2         Col3         Col4
  1     1016.46       991.29       960.21       708.38
  2      991.29       979.44       948.73       699.91
  3      960.21       948.73       940.21       693.63
  4      708.38       699.91       693.63      1079.38

                 Fit Statistics
      -2 Res Log Likelihood           384.7
      AIC (smaller is better)         398.7
      AICC (smaller is better)        401.5
      BIC (smaller is better)         404.1

        Null Model Likelihood Ratio Test
          DF    Chi-Square      Pr > ChiSq
           1         54.55          <.0001

         Type 3 Tests of Fixed Effects
               Num     Den
  Effect        DF      DF     F Value    Pr > F
  ia             1      12        0.65    0.4352
  sc             1      12        0.45    0.5145
  ia*sc          1      12        2.28    0.1567
  day            3    17.4        8.93    0.0008
  ia*day         3    17.4        0.53    0.6706
  sc*day         3    17.4        1.08    0.3841
  ia*sc*day      3    17.4        1.77    0.1899
```

TYPE=ANTE(1) does produce "Fit Statistics" values that are better than the TYPE=CS model. But, ANTE(1) requires estimation of 7 covariance parameters compared to the 2 required for CS. Let's see if TYPE=SP(EXP) can improve upon this model fit.

Step 6. Fine-Tune the Model Specification

6. a. i. Define & Fit Candidate Correlation Structures (cont'd)

Output Listing for Program 5.4a - TYPE=SP(EXP):

```
                     Dimensions
       Covariance Parameters          2

            Covariance Type is  SP(EXP)

     Estimated R Matrix for cow(ia*sc) 4 0 0
Row        Col1        Col2        Col3        Col4
  1     1033.64     1000.78      968.98      851.54
  2     1000.78     1033.64     1000.78      879.49
  3      968.98     1000.78     1033.64      908.36
  4      851.54      879.49      908.36     1033.64

                 Fit Statistics
      -2 Res Log Likelihood          407.3
      AIC (smaller is better)        411.3
      AICC (smaller is better)       411.5
      BIC (smaller is better)        412.8


        Null Model Likelihood Ratio Test
          DF    Chi-Square      Pr > ChiSq
           1         82.87          <.0001

         Type 3 Tests of Fixed Effects
               Num     Den
  Effect        DF      DF    F Value    Pr > F
  ia             1    12.3       0.59    0.4573
  sc             1    12.3       0.41    0.5349
  ia*sc          1    12.3       2.06    0.1759
  day            3    35.8       3.02    0.0424
  ia*day         3    35.8       0.19    0.9023
  sc*day         3    35.8       2.20    0.1046
  ia*sc*day      3    35.8       1.72    0.1811
```

Both TYPE=SP(EXP) and TYPE=ANTE(1) models indicate a substantial improvement in model fit when compared to TYPE=CS, and hence, all previously fit models. SP(EXP) yields larger fit statistics values than ANTE(1), but ANTE(1) requires estimation of 7 covariances while SP(EXP) estimates only two covariances.

Step 6. Fine-Tune the Model Specification

6. a. ii. Choosing a Suitable Covariance Structure

Since there are **no other** covariance structures that are **reasonable candidates** for this data, it is time to choose the most appropriate covariance structure.

The 'Fit Statistics' portion of the output listing from Program 5.4b provides a concise summary to facilitate the identification of the covariance structure that best fits the data set:

Fit Statistics for Candidate Covariance Structures

covtype	AIC	AICC	BIC	ResLog Like
ante1	398.7	401.5	404.1	384.7
cs	444.2	444.4	445.7	440.2
csh	448.1	449.6	452.0	438.1
sp_exp	411.3	411.5	412.8	407.3
un	401.3	407.2	409.0	381.3
vc	492.2	492.2	492.9	490.2

>>> An F-test CANNOT be used as a criterion for choosing the appropriate covariance structure. F-tests for fixed effects depend directly on the covariance structure, so it is essential to choose the best-fitting covariance structure for the data BEFORE conducting the F-tests.

This summary indicates UN, ANTE(1), and SP(EXP) all have comparably good values for the fit statistics. The statistical power and accuracy of F-test for the "within-cow" fixed-effect (i.e., day) relies upon both the accurate and parsimonious modeling of the covariance structure. The fewer covariance estimates required, the more degrees of freedom remain associated with the denominator of the F-tests. While there is no clear rule of thumb, selection of the model's covariance structure should strike a balance between "scientific expectation" and "parsimony".

>>> In the current example, UN provides a "saturated" covariance structure that can assist the researcher in confirming his/her expectation of the dependency structure present in the data. Viewing the **Estimated R** (above) for the UN, ANTE(1), and SP(EXP) models, **the researcher can practically interpret all of these** estimated covariance structures. The UN structure should seldom, if ever, be used because it is not parsimonious. Although **ANTE(1)** has the smallest AICC values, it estimates 7 covariances while **SP(EXP)** has a larger AICC value and estimates only 2 covariances. The degrees of freedom (DenDF) used to test for a significant Day effect are **17.4 for ANTE(1) and 35.8 for SP(EXP)**.

>>> **Based upon parsimony and visual inspection of the Estimated R, SP(EXP) is the covariance structure selected for this model.**

Step 6. Fine-Tune the Model Specification

6. a. ii. Choosing a Suitable Covariance Structure (cont'd)

Program 5.4b - SAS Macro to Facilitate Covariance Structure Selection

This SAS macro code runs the code in Program 5.4a for each candidate covariance structure and keeps only the information necessary to compare values of fit statistics. The covariance structure (other than TYPE=UN) with the smallest value of AIC, AICC, or BIC models your data the best of all the candidates considered. To run these SAS macros you need only type in the covariance TYPEs as parameters in the %id_cov( ) and %fitstats( ) macro calls.

Note: ***Program 5.4a is embedded in this macro.***

```
%MACRO id_cov (covtype, lbl) ;
        TITLE3 "Covariance Type is  &covtype" ;
        ODS OUTPUT  FITSTATISTICS= fit_&lbl ;
        ODS LISTING EXCLUDE ALL;
        PROC MIXED DATA = iasc ;
          CLASS cow ia sc day ;
          MODEL y = ia sc ia*sc
                            day day*ia day*sc day*ia*sc / DDFM= KR ;
          REPEATED / SUBJECT=cow(ia sc) TYPE=&covtype ;
        QUIT;
        DATA fit_&lbl;
         SET fit_&lbl;
                FORMAT covtype$ 7.;
                covtype="&lbl";
        RUN;
        ODS LISTING;
%MEND id_cov;

%id_cov( cs, cs );
%id_cov( vc, vc );
%id_cov( un, un );
%id_cov( csh, csh );
%id_cov( ante(1), ante1 );
%id_cov( sp(exp)(day), sp_exp );

DATA fitstats;
 SET fit_cs;
RUN;
%MACRO fitstats(ctype_lbl);
         DATA fitstats;
          SET fitstats fit_&ctype_lbl;
         IF MOD(_N_,4)=1 THEN stat_id='ResLogLike';
```

Step 6. Fine-Tune the Model Specification

6. a. ii. Choosing a Suitable Covariance Structure

Program 5.4b - **SAS Macro to Facilitate Covariance Structure Selection** (cont'd)

```
        IF MOD(_N_,4)=2 THEN stat_id='AIC     ';
        IF MOD(_N_,4)=3 THEN stat_id='AICC    ';
        IF MOD(_N_,4)=0 THEN stat_id='BIC     ';
      RUN;
%MEND fitstats;
%fitstats(vc);
%fitstats(un);
%fitstats(csh);
%fitstats(ante1);
%fitstats(sp_exp);
PROC SORT DATA=fitstats; BY covtype stat_id;
PROC TRANSPOSE DATA=fitstats OUT=tfits;
 VAR value;
 ID  stat_id;
 BY  covtype;
RUN;
DATA tfits; SET tfits; DROP _NAME_; RUN;
TITLE 'Fit Statistics for Candidate Covariance Structures';
PROC PRINT DATA=tfits; RUN;
```

Recall that bold caps indicate pre-defined statements, options or functions that SAS recognizes. Non-bold text is specific to your dataset. For the above TYPE= option in the REPEATED statement, you would choose one of the covariance structures listed in Appendix E.2 or in SAS/STAT documentation (referenced in Appendix C.7).

Program 5. 5 - ANOVA/Mean Comparisons - Correct Covariance Structure

```
PROC MIXED DATA = iasc ;
  CLASS cow ia sc day ;
  MODEL y = ia sc ia*sc
              day day*ia day*sc day*ia*sc / DDFM= KR OUTPRED=resids;
  REPEATED /SUB=cow(ia sc) TYPE=<    > R RCORR;
  LSMEANS   day   / DIFF;
  ODS OUTPUT DIFFS=PPP  LSMEANS=MMM;
  ODS LISTING EXCLUDE DIFFS  LSMEANS;
TITLE 'Covariance Structure is:  Name of the Covariance Structure';
RUN;
TITLE3 ' Means Comparisons for Significant Fixed Effects';
%INCLUDE 'c:\sas macros\ PDMIX800.SAS';
%PDMIX800(PPP,MMM,SORT=YES);
```

Step 6. Fine-Tune the Model Specification (cont'd)

Step 6. b. View the Covariance Matrix of the Fitted Model

The covariance structure of this data set was modeled using no RANDOM statement and the following REPEATED statement:

REPEATED / SUBJECT=cow(ia sc) **TYPE=SP(EXP)(**day**) R RCORR ;**

Since there is **no** effect specified in the **RANDOM statement**, by Eqn C.3.4 in Appendix C.3.4, $\mathbf{Var(Y_{80x1})_{80x80}} = \mathbf{R}_{80x80}$

\>>> As seen in Chapter 4, assigning an effect to the SUBJECT= option is equivalent to placing the same effect in a RANDOM statement, in the sense that, in either case, PROC MIXED models the levels of that effect as independent (not correlated) entities. In the current example, this is reasonable because there is no expectation that any of the 20 cows in the experiment should be dependent upon any other cow in the way she responds to the applied treatments. Therefore, **each element of $\mathbf{R}_{80x80}$ that represents the covariance between data from 2 different cows, will be zero**.

\>>> Regarding the covariance among measurements made on the same cow at different times, it is likely these measurements will exhibit some dependency on one another. In Step 5.b. above, Eqn 5.4 illustrates the general covariance structure, $\boldsymbol{R}_{k,}$ among the 4 time measurements taken on cow k (k=1,2,...,20). **$\mathbf{R}_{80x80}$ is created by placing these 20, 4x4 $\boldsymbol{R}_k$ matrices, along the diagonal**. The resulting $\mathbf{R}_{80x80}$ is said to be "block-diagonal". As illustrated in Step 6.a.i., the TYPE= option of the REPEATED statement was utilized to identify **SP(EXP)(**day**)** as the most appropriate covariance structure to describe the correlation observed among repeated time measures on the cows.

The covariance parameter estimates resulting from this model fit were:

Cov Parm	Subject	Estimate
$\sigma^2_{sp(exp)(Day)}$	**cow(ia*sc)**	**92.88**
σ^2_{Resid}		**1033.64**

As shown in Appendix E.2 for **TYPE=SP(EXP)(day)**, the covariance between measurements on 2 different days (i and j) on the same cow is a function of the distance apart in time of these measurements:

$$\mathbf{Cov\ (\ Day_i\ ,\ Day_j\ ) = \ \sigma^2_{Resid} \cdot(\ exp[-D(i\ ,\ j)\ /\sigma^2_{sp(Exp)(Day)}\ ]\ )}$$

where D(i, j) = | i - j | and i , j = 3, 6, 9, or 21

For each cow, (Note: With σ^2_{Resid} factored out, this is the 'correlation' matrix.)

$$\mathbf{R}_{k\ 4x4} = \mathbf{C}_{4x4} = 1033.64 \cdot \begin{array}{c} \\ \text{Day 3} \\ \text{Day 6} \\ \text{Day 9} \\ \text{Day 21} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Day 3} & \text{Day 6} & \text{Day 9} & \text{Day 21} \end{array} \\ \begin{bmatrix} 1 & e^{-3/92.88} & e^{-6/92.88} & e^{-18/92.88} \\ e^{-3/92.88} & 1 & e^{-3/92.88} & e^{-15/92.88} \\ e^{-6/92.88} & e^{-3/92.88} & 1 & e^{-12/92.88} \\ e^{-18/92.88} & e^{-15/92.88} & e^{-12/92.88} & 1 \end{bmatrix}_{4x4} \end{array}$$

$$= \quad 1033.64 \cdot \begin{bmatrix} 1 & .9682 & .9374 & .8238 \\ .9682 & 1 & .9682 & .8509 \\ .9374 & .9682 & 1 & .8788 \\ .8238 & .8509 & .8788 & 1 \end{bmatrix}_{4x4}$$

Therefore,

$$\mathbf{Var(Y_{80x1})_{80x80} = R_{80x80} =} \begin{bmatrix} C_{4x4} & 0_{4x4} & \cdots & 0_{4x4} \\ 0_{4x4} & C_{4x4} & \cdots & 0_{4x4} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{4x4} & 0_{4x4} & \cdots & C_{4x4} \end{bmatrix}_{80x80}$$

6. c. Check Model Diagnostics

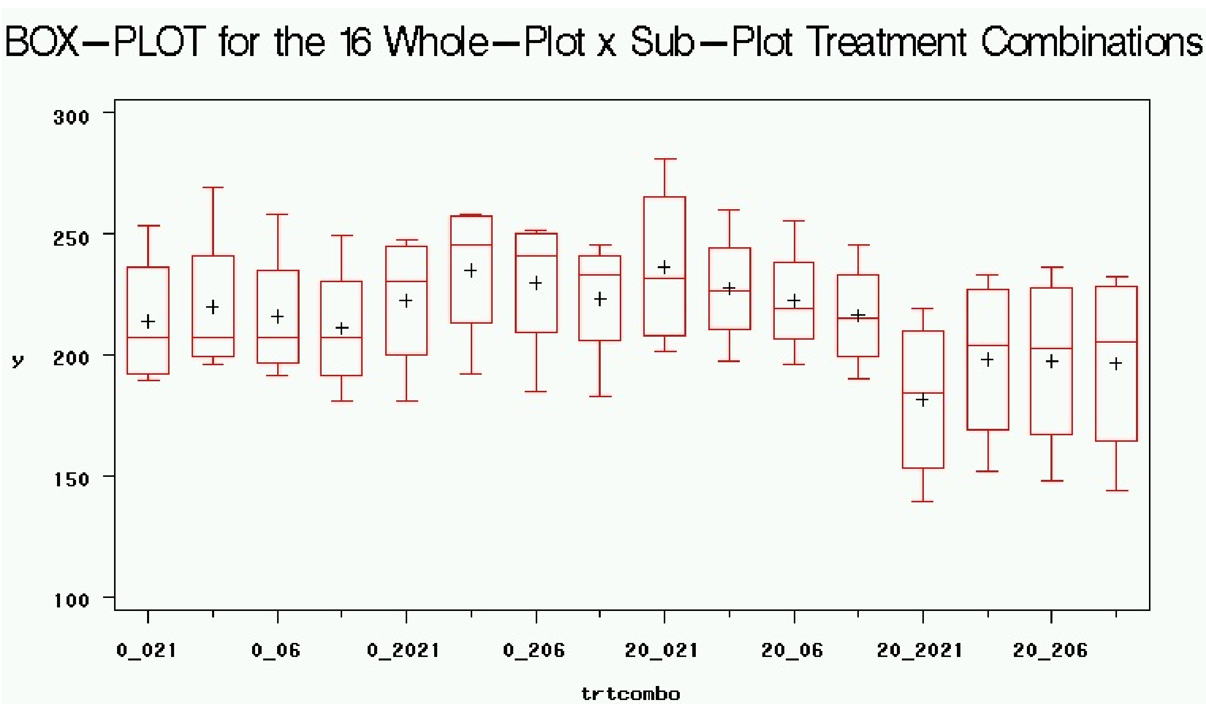
6. c. i. Check Normality of Error Variability

When there is a covariance structure to be fit to the data, this step should be conducted after the appropriate covariance structure has been identified and fit. SAS program code for checking normality using PROC UNIVARIATE has been given in all previous chapters so will not be repeated here.

6. c. ii. Check for Homogeneity of Treatment Variances

Before conducting the final analysis using Program 5.5 to obtain treatment means comparisons, it is still essential to check that the within-treatment variability is stable across all treatments. **Program 5.6** produces the below Box-plot.

```
PROC SORT        DATA= iasc  ;   BY trtcombo ;
PROC BOXPLOT DATA= iasc ;
         PLOT   y*trtcombo  ;
RUN;
```



The Box-plot indicates that within-treatment variability is relatively stable across treatments; especially since there were only 4 cows per treatment combination. Variance partitioning is unnecessary.

6. c. Check Model Diagnostics

6. c. iii. Will Transforming the Data Help? (Appendix B.2.3)

Data meet normality and homogeneity requirement of ANOVA.

6. c. iv. Calculate an R^2 Goodness-of-Fit Statistic for the Model (Appendix B.2.4)

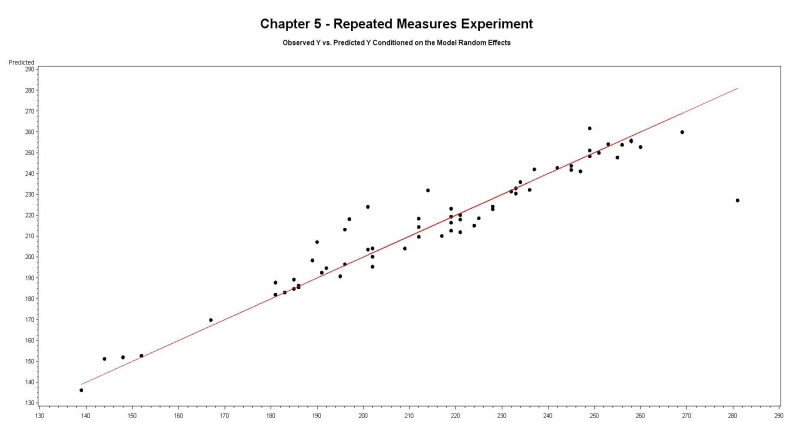
```
/* Calculate R-square of the fitted model.  */
title 'Chapter 5 Repeated Measures Experiment';
ods output FitStatistics=fitfull;
ods listing exclude all;
proc mixed covtest data=iasc method=ml;
  class cow ia sc day;
  model  y = ia sc ia*sc
              day day*ia day*sc day*ia*sc / ddfm=kr;
  repeated / subject=cow(ia sc) type=sp(exp)(day) r rcorr;
run;
ods output NObs=TotalN FitStatistics=fitIntOnly;
proc mixed covtest data=iasc method=ml;
  class cow ia sc day;
  model  y = ;
run;
/* Calculate # of observations for using to calculate R-square. */
data TotalN; set TotalN;
  where Label='Number of Observations Used';
  keep NObsUsed;
run;
data llfullmodel(rename=(Value=llfull));
 set fitfull;
   where substr(Descr,1,1)='-';
run;
data llfullmodel; set llfullmodel; keep llfull; run;
data llintonly(rename=(Value=llint));
 set fitintonly;
   where substr(Descr,1,1)='-';
run;
data llintonly; set llintonly; keep llint; run;
data TotalN; set TotalN; keep NObsUsed; run;
data rsq; merge llfullmodel llintonly TotalN;
  rSquare=1-exp((llfull-llint)/NObsUsed);
run;
ods listing;
title3 "Model's R-square is:";
proc print data=rsq noobs;
 var rsquare;
run;
```

The above SAS program calculates **the Model's R^2 = 0.86209**

6. c. Check Model Diagnostics

6. c. v. Calculate & Graph Model Predicted vs. Observed Values (Appendix B.2.5)

```
/* Plot the model predicted values vs. the observed original data
   to see how closely the graph resembles a 45-degree line
   through the origin.  */
data periods;
 set iasc;
     y=.;
    predtype='cond';
run;
data twice;
 set iasc periods;
run;
title 'Chapter 5 - Repeated Measures Experiment';
title3 'Observed Y vs. Predicted Y Conditioned on the Random Effects';
proc mixed data=twice;
  class cow ia sc day;
  model  y = ia sc ia*sc
                day day*ia day*sc day*ia*sc / ddfm=kr outpred=po;
   /* To obtain residuals conditioned on cows, need to specify 'cow'
   within treatment using the RANDOM statement. */
   random cow(ia sc);
   *repeated / subject=cow(ia sc) type=sp(exp)(day) r rcorr;
run;
data observed;
 set iasc;
run;
data cpred;
 set po;
  where predtype='cond';
  drop y;
run;
proc sort data=observed; by obs_id; run;
proc sort data=cpred;    by obs_id; run;
data cpred; merge observed cpred; by obs_id; run;
symbol1 i=none v=dot c=black;
symbol2 i=join v=none l=1 c=red;
proc gplot data=cpred;
  plot pred*y y*y / overlay;
run;
```



Step 7. Conduct Statistical Analysis

The 7 candidate models fit to the data on pages 5-8 through 5-13, indicate widely changing results regarding the significance of the "within-cow" (i.e, Day) fixed effect. This underscores the importance of modeling the covariance structure as accurately as possible. Below is the output listing for the mean comparison among DAY means resulting from Program 5.5 with TYPE=SP(EXP). All other output from Program 5.5 has already been shown above on page 5-13.

```
                              Covariance Type is  SP(EXP)

                     Means Comparisons for Significant Fixed Effects

--------------------- Effect=day A=LSD(.05) avgSD=6.58862 maxSD=9.4646 ----------------------

                                       Standard                          Pr >    Let
           Obs    day    Estimate       Error        DF     t Value       |t|    Grp
             1      3      220.00      8.0375      14.2       27.37    <.0001    A
             2      6      216.19      8.0375      14.2       26.90    <.0001    A
             3     21      213.50      8.0375      14.2       26.56    <.0001    AB
             4      9      211.75      8.0375      14.2       26.35    <.0001    B
```

The marginal (*P*=.0424) significance of the DAY effect suggests that 4 cows per treatment was a minimal amount of replication to achieve a 95% level of significance. **These results indicate the biological marker exhibits the largest reduction (i.e., signal of medical condition change) at 9 days**. A statistical power analysis of these data would likely recommend use of more than 4 cows per treatment in future studies.