Paper 178-2007

Repeated Measures Analysis with Clustered Subjects

Ramon C. Littell, University of Florida, Gainesville, FL

ABSTRACT

Repeated measures can occur on subjects that are nested within clusters. The clusters may be considered a random effect. Then the covariance structure must accommodate between-cluster variance as well as within-cluster covariance. There also may be higher levels of clustering. In addition, there may be other fixed effect factors that are crossed with some or all of the random effects. The covariance structure must be built into a statistical for the data and the statistical model must be translated into the language of computer software that has capabilities to make computations for statistical data analysis.

INTRODUCTION

The term "repeated measures" refers to data that are obtained on multiple occasions on the same subject. The subject could be a person, animal, plant, machine, or other unit from which data may be obtained. The "unit" is usually called a "subject" in statistical terminology, due to the human subject heritage of repeated measures analysis methods. Most commonly, repeated measures are taken over time, but they also could be obtained over space. More generally, repeated measures could refer to a multivariate response of similar data from a subject. In these situations, repeated measures can be considered another "factor" with levels given by points in time, space, or measurement scale. One of the primary difficulties in analyzing repeated measures data is determining an appropriate covariance structure for the data. The covariance structure must address covariance within the subjects on which repeated are obtained. Also, the covariance structure must incorporate variance and covariance due to random factors, such as clusters in which the subjects might be nested. In addition, variance and covariance due to experimental or sampling design must be built into the statistical model. This process usually falls into the realm of "hierarchical" modeling, because random effects tend to follow a nested classification.

In this paper we use an example from educational research to illustrate development of a statistical model. We also show how the statistical model can be implemented into the missed procedure of SAS®. Along the way, we emphasize that model development utilizes statistical analytic principles combined with subjective professional skills of statistical practice.

BUILDING A HIERARCHICAL LINEAR MODEL

Nested models are special cases of hierarchical models. (See Raudenbush and Bryk, 2002.) In nested models, there is a hierarchy of random effects corresponding to increasingly larger units. In the more general hierarchical models, there is also a nesting of units. In addition, there may be fixed effects that are crossed with the random effects. The hierarchy comes from first developing a model for the smallest sized units. Then, the smallest units are declared to be samples from populations of larger sized units, and random effects are assigned accordingly.

There are two steps in model development. First is the conceptualization step in which you think of the situation and develop the statistical model. Second is the programming step in which you translate the mathematical into a computer program that can provide data analysis according to the statistical model.

CONCEPTUALIZATION STEP: THE STATISTICAL MODEL

One of the most popular uses of hierarchical modeling is in educational research. Here is a common type of application. Students in a school district take a standardized exam and get scores on a quantitative and a qualitative scale. The students are in classes taught by teachers, so students may be considered "nested" within teachers. Likewise, teachers are nested within schools. In this example, we consider students and teachers to be nested random factors. Depending on the situation, the schools could be considered either fixed or random. In this application, we consider the schools fixed because we have data on all the school in the district, and we only wish to make inference about the schools in the district.

Since the scales on the exam were the same for all students, teachers, and schools, the scale factor is crossed with students within teachers, with teachers within schools, and with schools. "Scale" is called an observation-level factor because the levels (qual and quan) change within the two observations from each student. In addition, the students are classified according to a social-economic (SES) criterion with two levels (0 and 1). This factor is crossed with teachers and schools. "SES" is called a student-level factor because the level is SES is the same for a given student, but changes from student-to-student within the same teacher. The SAS data set for this example is named "score."

Hierarchical modeling starts with defining a model on the student level. This will be a model for the student data *conditional* on the teacher. Let y_{sm} be the score obtained by student s on scale m. We can write a model for the data for this student as

$$y_{sm} = \mu_m + e_{sm} \tag{1}$$

Where μ_m is the mean score for the "population" from which student s was obtained, and e_{sm} is the deviation for student k from the "population" mean. Let $m=1$ represent the qualitative scale, and *m*=2 represent the quantitative scale. Then the data for student s is the vector $y_s = (y_{s1}, y_{s2})'$. The error vector for this student is $e_s = (e_{s1}, e_{s2})'$. We can assume that e_s has covariance matrix $V(e_s) = R_s$.

The model equation can be written in terms of fixed effects for scale as

$$y_{sm} = \mu + \kappa_m + e_{sm} \tag{2}$$

where $\mu_m = \mu + \gamma_m$. Remember that this is a model for a student from a particular population. In this example, the population would correspond to the teacher who taught the student and the SES group of the student. We can generalize the model to include the effects of SES criterion to be

$$y_{ksm} = \mu + \delta_k + \kappa_m + (\delta\kappa)_{km} + e_{ksm} \tag{3}$$

where β_k is the fixed effect of SES criterion k and $(\beta\gamma)_{km}$ is the interaction between SES criterion k and test scale m.

Don't forget that this is the model for a student conditional on a teacher. Assume student s was taught by teacher j. Let y_{jksm} be the score obtained by student s on scale m. We can add random effect of teacher to the model. In this example, we allow the possibility that the effect of teacher depends on the scale of the exam. Some teachers are better at teaching qualitative subjects and others are better at teaching quantitative subjects. The expanded model could be written

$$y_{jksm} = \mu + \delta_k + \kappa_m + (\delta\kappa)_{km} + d_{jm} + e_{jksm} \tag{4}$$

where d_{jm} is the random effect for teacher j on scale m. We can assume that the vector $\mathbf{d}_j = (d_{j1}, d_{j2})'$ has covariance matrix $V(\mathbf{d}_j) = G_j$. Finally, we can expand the model to contain effects of school. We include terms for interaction between school and scale and school and SES criterion. The model becomes

$$y_{ijksm} = \mu + a_i + \delta_k + (a\delta)_{ik} + \kappa_m + (\alpha\kappa)_{im} + (\delta\kappa)_{km} + d_{jm} + e_{ijksm} \tag{5}$$

where α_i, $(\alpha\beta)_{ik}$, and $(\alpha\gamma)_{im}$ are effects of school, school by SES criterion interaction, and school by scale interaction, respectively.

Assume that there I schools, J teachers, and S students. For simplicity, assume also that the schools, teachers, and students have unique identification codes. In the linear mixed model notation,

$$Y = X\beta + ZU + e \tag{6}$$

The fixed effect vector β contains the parameters $\mu, a_i, \delta_k, (a\delta)_{ik}, \kappa_m, (\alpha\kappa)_{im}, (\delta\kappa)_{km}$, U contains the teacher random effects d_{jm}, and e contains the student random effects e_{jksm}. The matrix $G = V(U)$ is block-diagonal, with the 2x2 sub-matrices $G_1, G_2, \ldots, G_J$ on the diagonal:

$$G = \begin{pmatrix} G_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & G_J \end{pmatrix} \text{ and } G_j = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \qquad (7)$$

The formulation (8.7) declares that G_j is the same for each teacher. It says that the variance of the teacher effect is γ_{11} for the qualitative scale (1) and γ_{22} for the quantitative scale (2), and the covariance is γ_{12} between the qualitative and quantitative scales

The matrix V(**e**) = **R** is block-diagonal, with the 2x2 sub-matrices R_1, R_2, … , R_S on the diagonal:

$$R = \begin{pmatrix} R_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_S \end{pmatrix} \text{ and } R_s = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} \qquad (8)$$

Similar to the situation with teachers, formulation (8) declares that R_s is the same for each student.

The overall variance matrix for the model (6) is

$$V = V(Y) = ZGZ' + R \qquad (9)$$

The matrix V is block-diagonal of form

$$V = \begin{pmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_J \end{pmatrix} \qquad (10)$$

Where V_j is the variance matrix for the data obtained from the students taught by teacher *j*. The matrix V_j is given by

$$V_j = Z_j G_j Z_j^{'} + J_j \otimes R_s \qquad (11)$$

Where J_j is the square matrix containing 1 in every position of dimension equal to number of students in the class taught by teacher j.

The matrices G_j and R_s can be structured. For example, if $\lambda_{11} = \lambda_{22} = \lambda$ and $\gamma_{12} = \rho\lambda$, then R_s has *compound symmetric* structure. If there are no functional relationships between the parameters, then the matrix is *unstructured*. The condition $\lambda_{11} = \lambda_{22}$ would say that the variance of student scores (conditional on teacher) is the same on the qualitative scale as on the quantitative scale. In general, we want the simplest model that adequately represents the data. So, we want to use cs structure if it is justified; that is, to assume equal variances for the qualitative and quantitative scales. Making this decision is more easily said that done. Indeed, model selection is perhaps the most difficult part of good statistical analysis.

PROGRAMMING STEP: TRANSLATING THE STATISTICAL MODEL INTO SAS CODE

Model (6) is the basic model implemented by the mixed procedure in SAS®. The most challenging part is to decide on an appropriate variance matrix **V** and to construct it using the capabilities of `proc mixed`. In this section we discuss how to implement model (6) for the scores data into `proc mixed`. See Littell, Milliken, Stroup, Wolfinger and Schabenberger (2006) for general information about `proc mixed`. See Wolfinger (1996) and Verbeke and Melonberghs (2000) for other considerations in modeling random repeated measures and longitudinal data.

There are usually several ways to implement a model in a given application. That is true in the present example. We shall take the most direct approach, if not the most computationally efficient. We shall model the teacher random effects using the random statement and the student random effects using the repeated statement. In fact, the random statement could be used for both teacher and student. The basic statements will be of the form

```
proc mixed data=score;
  class school teacher student ses scale;
  model score=fixed effects;
  random random effects for teacher;
  repeated covariance structure for students;
run;
```

The `model` statement is relatively straightforward. It follows directly from the fixed effects is model equation (5):

```
model score=school ses school*ses scale school*scale ses*scale;
```

Options can be added to the model statement, for example, to select the method of computing degrees of freedom. In this application, we shall use the option `ddfm=kw` to select the kenward-roger method.

In most applications, the random statement is used to define simple random effects for a factor that are assumed to be normally distributed with variance, say γ , by using the statement

```
random teacher;
```

In that case, the **G** matrix of (7) would be diagonal with γ in each diagonal position. But in this example there are two random effects for teachers, one for the qual scale and one for the quan scale, making the **G** matrix block-diagonal with $G_j = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$ in each block position. This is what we must specify in the `random` statement. Here is code to do that:

```
random scale / sub=teacher(school) type=un;
```

This statement basically says there are two variables (qual and quan) represented by "`scale`" that occur for each teacher. We specify "teacher" as the unit on which the variables are defined by declaring "teacher" to be the "subject." That is the purpose of `sub=teacher(school)`. The "type" of covariance structure is defined by the `type=un` designation.

The student random effects can be modeled directly with a repeated statement.

```
repeated scale/sub=student(teacher school) type=un;
```

In fact, it is not necessary to designate "scale" in the repeated statement because that is the only factor with levels changing within students, and each student has responses for both qual and quan.

The basic statements will be

```
proc mixed data=score;
  class school teacher student ses scale;
  model score=school ses school*ses scale school*scale ses*scale;
  random scale / sub=teacher(school) type=un;
  repeated scale/sub=student(teacher school) type=un;
run;
```

In the next section we use these statements to select a covariance structure and then add optional statements to estimate means and make comparisons of means for appropriate combinations of ses and scale.

ANALYSIS STORY

ANALYSIS FOR SCALES SEPARATE

It is often useful to analyze subsets of the data separately in the development of an overall model. In the present case, it is informative to perform separate analyses for the qualitative and quantitative scales before combining them into an overall model. This allows us to examine directly the variance due to teacher and student of the score for each scale. That will help us to formulate the G and R matrices. We can run `proc mixed` for each scale using the statements

```
proc sort data=scores; by scale;
run;
proc mixed data=score; by scale;
  class school teacher ses;
  model score = school ses school*ses/ddfm=kr;
  random teacher(school);
run;
```

Here are some of the results from the output:

First are the estimates of variance and covariance parameters for qual and quan.

scale=qual

Covariance Parameter Estimates				
Cov Parm	**Estimate**	**Standard Error**	**Z Value**	**Pr Z**
teacher	2.9666	2.5503	1.16	0.1224
Residual	133.32	6.5347	20.40	<.0001

scale=quan

Covariance Parameter Estimates				
Cov Parm	**Estimate**	**Standard Error**	**Z Value**	**Pr Z**
teacher	8.0573	3.5037	2.30	0.0107
Residual	105.23	5.1623	20.38	<.0001

In summary:

Scale = qual: Variance Component estimates: $\hat{\gamma}_{qual} = 2.97$ $\hat{\lambda}_{qual} = 133.32$

Scale = quan: Variance Component estimates: $\hat{\gamma}_{quan} = 8.06$ $\hat{\lambda}_{quan} = 105.23$

These results suggest quite different variances for the two scales on the teacher level, but similar variances on the student level.

Here are results from the tests of fixed effects:

scale=qual

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
school	13	47.2	1.89	0.0555
ses	1	861	23.90	<.0001
school*ses	13	856	1.94	0.0232

scale=quan

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
school	13	43.6	1.78	0.0787
ses	1	856	10.61	0.0012
school*ses	13	850	0.97	0.4813

These results show different results in the school*ses interaction for the two scales, suggesting that ses effects differ between schools in the qual scale but not in the quan scale.

ANALYSIS FOR SCALES COMBINED

Now we use proc mixed to obtain an analysis for the combined data. Here are the statements:

```
proc mixed data=score;
  class school teacher student ses scale;
  model score=school ses school*ses scale school*scale ses*scale;
  random scale / sub=teacher(school) type=un;
  repeated scale/sub=student(teacher school) type=un;
run;
```

Here are results for the covariance parameter estimates:

Covariance Parameter Estimates		
Cov Parm	**Subject**	**Estimate**
UN(1,1)	teacher(school)	3.3279
UN(2,1)	teacher(school)	4.5153
UN(2,2)	teacher(school)	8.2399
UN(1,1)	student	133.46
UN(2,1)	student	22.8222
UN(2,2)	student	105.36

The variance parameter estimates are:

Teacher-level variance parameter estimates: $\hat{\gamma}_{qual} = 3.33$ $\hat{\gamma}_{qual,quan} = 4.51$ $\hat{\gamma}_{quan} = 8.24$

Student -level variance parameter estimates: $\hat{\lambda}_{qual} = 133.46$ $\hat{\lambda}_{qual,quan} = 22.82$ $\hat{\lambda}_{quan} = 105.36$

These estimates are consistent with those from the separate analyses: teacher-level variances are different between qual and quan, but student level variances are similar.

This table gives basic information about model fit. It shows (-2) residual log likelihood (-2 Res Log L) and three indices (AIC, AICC, and BIC) that are derived from the -2 Log L, which make various adjustments related to the amount of data and numbers of parameters in the covariance structure.

Fit Statistics	
-2 Res Log Likelihood	13381.6
AIC (smaller is better)	13393.6
AICC (smaller is better)	13393.7
BIC (smaller is better)	13385.8

We run code to try all four combinations of covariance structure (un and cs) on the two levels (teacher and students). Results are summarized here:

teacher	student	-2Res Log L	AIC	BIC
un	un	13381.6	13393.6	13385.5
cs	un	13383.3	13383.3	13386.8
un	cs	13393.6	13404.6	13397.1
cs	cs	13394.6	13402.6	13397.3

You see that the results for un-un come from the table above.

It is rather surprising that the results for un-un are very close to those for cs-un. This suggests that cs might be a better choice for the teacher-level covariance structure than un. Indeed, that is what the AIC criterion suggests, contrary to the perception we obtained from the separate analyses and the un-un parameter estimates. It is also surprising that the un-cs results differ so much from un-un because our previous impression was that cs might be suitable for student-level covariance structure. For what its worth, selection based on BIC goes to un-un. In addition, preliminary considerations (not reported here) using analysis of variance suggested the un structure for teachers. For these reasons, we choose the un-un structure for the covariance matrix in the model from which we make inference about fixed effects.

The tests of fixed effects show interaction between ses and scale. Also, there is interaction between school and each of ses and scale. However, we shall proceed with inference about effects of ses and scale averaged over schools.

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
school	13	42.9	1.78	0.0779
ses	1	862	28.63	<.0001
school*ses	13	850	1.52	0.1043
scale	1	40.2	56.00	<.0001
school*scale	13	35.1	2.23	0.0296
ses*scale	1	873	4.38	0.0367

We add an `lsmeans` statement to make inference about ses and scale.

```
proc mixed data=score;
  class school teacher student ses scale;
  model score=school ses school*ses scale school*scale ses*scale;
  random scale / sub=teacher(school) type=un;
  repeated scale/sub=student(teacher school) type=un;
  lsmeans ses*scale / pdiff;
run;
```

Output from the `lsmeans` statement is:

Least Squares Means							
Effect	**scale**	**ses**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
ses*scale	qual	0	35.9238	0.7931	170	45.30	<.0001
ses*scale	quan	0	32.8110	0.8106	112	40.48	<.0001
ses*scale	qual	1	42.1397	0.8897	256	47.37	<.0001
ses*scale	quan	1	36.7242	0.9166	176	40.07	<.0001

Differences of Least Squares Means									
Effect	**scale**	**ses**	**_scale**	**_ses**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
ses*scale	qual	0	quan	0	3.1129	0.8403	169	3.70	0.0003
ses*scale	qual	0	qual	1	-6.2158	1.1330	1091	-5.49	<.0001
ses*scale	qual	0	quan	1	-0.8004	1.1087	428	-0.72	0.4707
ses*scale	quan	0	qual	1	-9.3287	1.1008	413	-8.47	<.0001
ses*scale	quan	0	quan	1	-3.9132	1.0552	1017	-3.71	0.0002
ses*scale	qual	1	quan	1	5.4154	0.7410	107	7.31	<.0001

The tables of lsmeans and their differences show that qual scores are higher that quan scores in ses=0. The difference is 3.1 with standard error 0.84. The difference between qual and quan in ses=1 is 5.42 with standard error 0.74. Hence the interaction between scale and ses due to the larger difference between qual and quan scores in ses 1 than ses 0. The difference between qual scores for ses=0 is -6.18 and the difference between quan scores for ses=1 is -3.91, again explaining the interaction.

COMPARISON OF FIXED EFFECT INFERENCE USING A DIFFERENT COVARIANCE STRUCTURE

It is interesting to see how the fixed effect inference depends on the choice of covariance structure. We change to cs for teachers and run the statements

```
proc mixed data=score;
  class school teacher student ses scale;
  model score=school ses school*ses scale school*scale ses*scale;
  random scale / sub=teacher(school) type=cs;
  repeated scale/sub=student(teacher school) type=un;
  lsmeans ses*scale / pdiff;
run;
```

Here is the table for tests of fixed effects:

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
school	13	66	1.75	0.0700
ses	1	1671	28.53	<.0001
school*ses	13	1671	1.52	0.1041
scale	1	66	56.26	<.0001
school*scale	13	66	2.25	0.0164
ses*scale	1	1671	4.27	0.0389

Comparison with the table using un structure shows there is very little change in the test of fixed effects.

The table of lsmeans is:

Least Squares Means							
Effect	**scale**	**ses**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
ses*scale	qual	0	35.9381	0.8326	1671	43.16	<.0001
ses*scale	quan	0	32.7867	0.7768	1671	42.21	<.0001
ses*scale	qual	1	42.1198	0.9242	1671	45.57	<.0001
ses*scale	quan	1	36.6968	0.8867	1671	41.38	<.0001

These results are also very similar to the corresponding results using un for teacher covariance.

The differences of lsmeans are:

Differences of Least Squares Means									
Effect	**scale**	**ses**	**_scale**	**_ses**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
ses*scale	qual	0	quan	0	3.1515	0.8410	1671	3.75	0.0002
ses*scale	qual	0	qual	1	-6.1817	1.1310	1671	-5.47	<.0001
ses*scale	qual	0	quan	1	-0.7587	1.1091	1671	-0.68	0.4940
ses*scale	quan	0	qual	1	-9.3332	1.0993	1671	-8.49	<.0001
ses*scale	quan	0	quan	1	-3.9101	1.0536	1671	-3.71	0.0002
ses*scale	qual	1	quan	1	5.4231	0.7418	1671	7.31	<.0001

Again, results are not much different from those using un. So you might wonder whether it really matters what covariance structure is used. Sometimes it does and sometimes it does not. In this example it does not matter very much. But that is because the greatest change was in the teacher variance component, and those variance are small compared to the student variances. It the reverse had been true, you would see a big difference in the standard errors of the lsmeans for the two scales.

OTHER SITUATIONS

In the previous sections the “repeated measures” were simple the bivariate responses from the qualitative and quantitative scales. If the repeated measures are responses over time, then other methods might be more appropriate. Suppose that instead of the two response scales, each student yielded measurements at five points in time. Denote these as times 1 through 5. Then, instead of the factor “scale” we would have the factor “time.”

TIME AS A QUALITATIVE FACTOR

Sometimes it is suitable to consider time a factor with discrete qualitative levels. This may be the case when discrete changes from one time to another are more relevant than the overall trend over time. Often there is no readily identifiable trend that can be modeled with mathematical functions. In this situation the methods of analysis are a direct extension from those discussed in previous sections. SAS code might be

```
proc mixed data=score;
  class school teacher student ses time;
  model score=school ses school*ses time school*time ses*time;
  random teacher(school);
  repeated time/sub=student(teacher school) type=un;
  lsmeans ses*time / pdiff;
run;
```

This code takes teacher as a factor with random effects that does not define covariance between times at the teacher level. This concept is seldom addressed, but could be an issue as surely as it was in the discussions of the previous sections. If it is addressed, it could lead to a computational problem due to the numerous parameters that would be required in the covariance matrices.

TIME AS A QUANTITATIVE FACTOR

If trends over time are of interest, then an analysis should be used that treats time as a regression variable. In such cases a preliminary analysis treating time as a qualitative variable is usually advisable to identify covariance structure. (See Littell, Pendergast and Natarajan, 2000). After covariance structure is selected, the fixed effects of time can be modeled as curves over time. Doing this using proc mixed requires manipulation of variable names to essentially define covariance structure using discrete levels of time and fixed effects using time a regression variable. Then you need two time variables so you can treat one as qualitative and one as quantitative. You could use code like this:

```
data score2; set score; t=time;
run;
proc mixed data=score2;
  class school teacher student ses time;
  model score=school ses school*ses t school*t ses*t;
  random teacher(school);
  repeated time/sub=student(teacher school) type=un;
run;
```

In place of the `lsmeans` statement you could use `estimate` statements to estimate slopes, differences between slopes, etc. Also, when treating time a quantitiave variable, you might want to utilize time series covariance structures, such as autoregressive. You would replace the `repeated` statement as follows:

```
  random student (teacher school);
  repeated time/sub=student(teacher school) type=ar(1);
run;
```

This code adds a random effect for student for the between-student structure, and autoregressive covariance within students. A word of warning: You would be getting into deep water with the modeling and would likely encounter computational issues.

Another approach is to use the so-called random coefficient regression approach. This approach does not explicitly model the covariance structure within students, but instead simply fits “random” regression models for each student. This can be done with this code:

```
proc mixed data=score2;
  class school teacher student ses;
  model score=school ses school*ses t school*t ses*t;
  random teacher(school);
  random intercept t/sub=student(teacher school) type=un;
run;
```

Notice that time is not used as a classification variable at all.

SUMMARY AND CONCLUSIONS

In this paper we discuss building a statistical model for repeated measures for clustered subjects using hierarchical modeling. The statistical model is implemented in `proc mixed` using the `random` and `repeated` statements. Choosing a covariance structure is discussed, with a final decision based on several considerations, rather than a single index or test of hypothesis. Then inference about fixed effects is obtained.

REFERENCES

Littell, R.C., J. Pendergast, and R. Natarajan. 2000. "Modelling Covariance Structure in the Analysis of Repeated Measures Data." *Statistics in Medicine 19:1793-1819.*

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2006. *SAS for Mixed Models, 2nd ed.* SAS Press, Cary, NC.

Raudenbush, S.W., and A.S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage Publications, Newbury Park, CA.

Verbeke, G. and G. Molenberghs. 2000. *Linear Mixed Models for longitudinal Data.* Springer, New York.

Wolfinger, R.D. 1996. "Heterogeneous Variance Covariance Structure for Repeated Measures." *J. of Agricultural, Biological, and Environmental Statistics 1(2): 2005-230.*

ACKNOWLEDGMENTS

Thanks to Jennifer Waller and Maribeth Johnson for their assistance with preparation of the manuscript.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ramon C. Littell
University of Florida
Box 118545
Gainesville, Florida 32611-8545
E-mail: Littell@ufl.edu