

Paper 188-29

Repeated Measures Modeling With PROC MIXED

E. Barry Moser, Louisiana State University, Baton Rouge, LA

ABSTRACT

PROC MIXED provides a very flexible environment in which to model many types of repeated measures data, whether repeated in time, space, or both. Correlations among measurements made on the same subject or experimental unit can be modeled using random effects, random regression coefficients, and through the specification of a covariance structure. PROC MIXED provides a large variety of useful covariance structures for modeling covariation in both time and space, including discrete and continuous increments of time and space. MANOVA tests are available for some model specifications, and degrees of freedom adjustments are available to provide better approximations to the distributions of the test statistics than for standard between- or within-subject methods. The %GLIMMIX macro, available in the SAS/STAT sample library, extends the mixed model technology of PROC MIXED to generalized linear mixed models, while the %NLINMIX macro, also available in the SAS/STAT sample library, provides a similar framework for non-linear mixed models. Likelihood and information criteria are available to aid in the selection of a model when the model structure is not known *a priori*.

INTRODUCTION

Repeated measures data are encountered in a wide variety of disciplines including business, behavioral science, agriculture, ecology, and geology. What generally distinguishes repeated measures data from time series data is that multiple subjects are involved, and the number of measurements per subject is generally not very large. The amount of literature on repeated measures analysis is now quite large (see the references section for some recent books on the topic), partly as a result of advances in statistical computing that permit the use of methods that are often quite computational in detail. The reason for the large number of publications and software solutions is that there is no simple methodology that accommodates all facets of repeated measures experiments and surveys. For example, repeated measures experiments may measure each subject at each of a designated set of equally spaced times, while another experiment may result in the number of measurements per subject as a random variable with measurements made at irregularly spaced times. Other experiments may have drop outs while the experiment progresses. The drop outs may be random, or may be a result of the treatment that the subject receives. In some experiments, the repeated measures factor can be randomized, while in others, the repeated measures factor is time or space and cannot be randomized. In addition, the response variable or variables may be continuous, discrete, or a combination of the two, leading to different possibilities for making inference from the experiment or survey.

PROC MIXED in the SAS System provides a very flexible modeling environment for handling a variety of repeated measures problems. Random effects can be used to build hierarchical models correlating measurements made on the same level of a random factor, including subject-specific regression models, while a variety of covariance and correlation structures can be specified for residuals. The random effects and covariance structures are specified in the RANDOM and REPEATED statements, respectively.

RANDOM STATEMENT

The RANDOM statement specifies the random effect terms that are to be included in the model, along with a covariance structure (TYPE= option) to specify how the random effects are related to each other. More than one RANDOM statement can be used to define a PROC MIXED model. This is useful when you have some random effects that are correlated, while there are others that need to be independent of those, as in hierarchical models. The SUBJECT= option can be used to specify different hierarchies.

REPEATED STATEMENT

The REPEATED statement controls the covariance structure imposed upon the residuals or errors. In procedures such as GLM and REG, the errors are assumed to be independent, while PROC MIXED has a rich variety of structures to specify relationships among the errors. In repeated measures models the SUBJECT= optional statement parameter is used to define which observations belong to the same subject, and which belong to different subjects, where different subjects are independent. The TYPE= optional statement parameter specifies the model for the covariance structure of the errors. The GROUP= optional statement parameter permits different levels of the GROUP effect to have different structure parameters, though the structure TYPE remains the same. Only one REPEATED statement is permitted in a PROC MIXED model.

RANDOM EFFECTS

Example 1 is taken from Lindsey (1993:86) and concerns the visual acuity of eyes and lens strength. Response times in msec of seven subjects were measured after a light is shown into each eye under each of 4 lens powers, 6/06, 6/18, 6/36, and 6/60, where the power is the perception of the object by the eye at 6 feet when it is actually positioned at 6, 18, 36, or 60 feet, respectively. The question of interest is whether response time varies with lens strength. Assume that the 8 lens/eye combinations are completely randomized separately for each subject, and that there is no

carryover effect. A carryover effect would be present if the response time of a measurement depended upon the treatment previously applied.

PAIRED MEASUREMENTS

If we were to analyze the data only for the 6/06 lens power with an interest in comparing the left eye with the right eye, then we would have 2 measurements made on each subject – one on the left eye and one on the right eye.

Measurements made on the same subject are likely to be correlated, while it seems reasonable to believe that the response observed on one subject is independent of that observed on another subject. A typical analysis of this data would be with a paired t-test (or a non-parametric analog). In the paired t-test, we are estimating the population mean difference of left- and right-eye responses, $\mu_L - \mu_R$, using the sample mean difference, $\bar{Y}_L - \bar{Y}_R$. The variance of this difference of means is

$$\text{Var}(\bar{Y}_L - \bar{Y}_R) = \sigma_L^2 + \sigma_R^2 - 2\sigma_{LR} \quad (1)$$

where σ_L^2 , σ_R^2 , and σ_{LR} are the variances and covariance of the left and right eyes. Unlike the two-independent samples problem, the covariance term is present and measures the linear dependence between the left and right-eye measurements. An experimental design model that results in this same variance of the difference of means is the randomized block design (RBD) model. This model can be written as

$$Y_{ij} = \mu + \tau_i + B_j + \varepsilon_{ij} \quad (2)$$

where Y_{ij} is the response measured on the i^{th} eye, $i=[\text{Left}, \text{Right}]$, on the j^{th} subject, μ is the overall mean, τ is the mean effect of eye, B is the random subject effect, ε is the error, and $B_j \sim N(0, \sigma_B^2)$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, and B_j and ε_{ij} are independent. This model can be fitted in a straightforward way using PROC MIXED. The MODEL statement will be used to specify the τ_i , while the RANDOM statement will be used to specify that the B_j are to be included and used to estimate σ_B^2 . The model defaults to include an intercept term, μ , and errors ε_{ij} .

```
Proc Mixed Data=Vision;
  Where LensStrength="6/06";
  Class Subject Eye;
  Model ResponseTime = Eye;
  Random Subject / V VCorr;
  LSMeans Eye;
Run;
```

The random effects are estimated to be 3.9048 and 32.9762 for the subject and error, respectively. The V and VCORR options of the RANDOM statement produce tables showing the estimated covariance and correlation matrices of the responses. These matrices for the first two subjects only are given in the tables below.

Table 1. Covariance matrix for subjects 1 and 2.

Row	Col1	Col2	Col3	Col4
1	36.881	3.905		
2	3.905	36.881		
3			36.881	3.905
4			3.905	36.881

Table 2. Correlation matrix for subjects 1 and 2.

Row	Col1	Col2	Col3	Col4
1	1.000	0.106		
2	0.106	1.000		
3			1.000	0.106
4			0.106	1.000

In these tables Col1 and Col2 refer to measurements made on subject 1, while Col3 and Col4 refer to the measurements made on subject 2. Further Col1 and Col3 correspond to the left eye and Col2 and Col4 to the right eye. We observe here that measurements made on the same subject are allowed to be correlated ($r=0.106$), while measurements between different subjects are independent (absence of covariances and correlations). However, we note that this model requires that the variances of the left and right eyes be equivalent, which does not quite match the variance of the differences of the mean model proposed earlier. To match that formula for the variance of the difference of the means, we'll need to permit the variance associated with the left eye measurements to differ from the variance associated with the right eye measurements. The following PROC MIXED code can accomplish this.

```
Proc Mixed Data=Vision;
  Where LensStrength="6/06";
  Class Subject Eye;
  Model ResponseTime = Eye;
  Random Subject / V VCorr;
  Repeated / Group=Eye;
  LSMeans Eye;
Run;
```

You might notice that the observed test statistic and the variance of the difference of the means is equivalent for these two analyses. Because there are two balanced groups (eyes), the average of the two residual variances in the second analysis is equivalent to the residual variance in the first model leading to these equivalences. As we build more complex models, however, this may not generalize.

BLOCKS OF MEASUREMENTS

Now let's look at the lens strength aspect of these data and focus on using just the left eye measurements. Now we have 4 measurements made on each subject. Let's fit the RBD model with homogeneous variance to these data.

```
Proc Mixed Data=Vision;
  Where Eye="Left ";
  Class Subject LensStrength;
  Model ResponseTime = LensStrength;
  Random Subject / V VCorr;
  LSMeans LensStrength / PDiff;
Run;
```

Table 3. Estimated covariance matrix of responses on the same subject for the 4 lens strengths, with lens powers 6/06, 6/18, 6/36, and 6/60, respectively.

Row	Col1	Col2	Col3	Col4
1	28.869	5.397	5.397	5.397
2	5.397	28.869	5.397	5.397
3	5.397	5.397	28.869	5.397
4	5.397	5.397	5.397	28.869

Table 4. Estimated correlation matrix of responses on the same subject for the 4 lens strengths, with lens powers 6/06, 6/18, 6/36, and 6/60, respectively.

Row	Col1	Col2	Col3	Col4
1	1.000	0.187	0.187	0.187
2	0.187	1.000	0.187	0.187
3	0.187	0.187	1.000	0.187
4	0.187	0.187	0.187	1.000

The covariance and correlation matrices are given in Tables 3 and 4, and note that the correlation between any pair of lens powers is constrained to be the same by using this RBD model. If we are able to randomize the order that each subject receives the lens power treatments, then on average, equi-correlated responses would be expected.

Next, consider permitting the variances of the responses associated with the different lens strengths to differ by using the GROUP= option on the REPEATED statement.

```
Proc Mixed Data=Vision;
  Where Eye="Left ";
  Class Subject LensStrength;
  Model ResponseTime = LensStrength / DDFM=KenwardRoger;
  Random Subject / V VCorr;
  Repeated / Group=LensStrength;
  LSMeans LensStrength / PDiff;
Run;
```

When we examine the covariance and correlation matrices of responses, we find that the responses are no longer equi-correlated. Further, we observe that one estimated variance is almost an order of magnitude greater than another. This is due to an outlying observation that could be excluded, and certainly its effects on the inferences should be investigated before drawing conclusions.

A major difference between this and the previous model is that the estimated variances of the differences of the least squares means now depend upon which pair of means is being compared (Table 7). Because of the unequal variances, the Kenward-Roger approximation to the degrees of freedom for the reference distribution is recommended (see Schabenberger and Pierce 2002: 430-431, 438-446).

Table 5. Estimated covariance matrix of responses on the same subject for the 4 lens strengths, with lens powers 6/06, 6/18, 6/36, and 6/60, respectively, for the heterogeneous variance model.

Row	Col1	Col2	Col3	Col4
1	13.248	8.192	8.192	8.192
2	8.192	10.800	8.192	8.192
3	8.192	8.192	82.902	8.192
4	8.192	8.192	8.192	17.111

Table 6. Estimated correlation matrix of responses on the same subject for the 4 lens strengths, with lens powers 6/06, 6/18, 6/36, and 6/60, respectively, for the heterogeneous variance model.

Row	Col1	Col2	Col3	Col4
1	1.000	0.685	0.247	0.544
2	0.685	1.000	0.274	0.603
3	0.247	0.274	1.000	0.218
4	0.544	0.603	0.218	1.000

Table 7. Differences of Least Squares Means for the heterogeneous variance RBD model.

Effect	Lens Strength	_Lens Strength	Estimate	Standard Error	DF	t Value	Pr > t
LensStrength	6/06	6/18	-0.714	1.046	4.67	-0.68	0.5272
LensStrength	6/06	6/36	2.714	3.376	6.56	0.80	0.4495
LensStrength	6/06	6/60	-3.857	1.413	7.97	-2.73	0.0259
LensStrength	6/18	6/36	3.429	3.324	5.99	1.03	0.3421
LensStrength	6/18	6/60	-3.143	1.283	6.99	-2.45	0.0442
LensStrength	6/36	6/60	-6.571	3.456	6.96	-1.90	0.0993

CORRELATED ERRORS

Let's now consider an alternative strategy for modeling the data of example 1. Again, we'll focus first on the paired data of lens power 6/06. Since two measurements are recorded on each subject, the set of responses on a subject form a multivariate response. A standard technique for asking questions concerning linear combinations of the responses (here, differences of left versus right eye means), is profile analysis (see e.g., Johnson and Wichern 2002). This type of analysis can be conducted readily using PROC GLM and its REPEATED statement, but PROC MIXED can also perform this analysis using a correlated errors model. The correlated errors model can be written as

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (3)$$

but now where

$$\text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = \text{Cov}(Y_{ij}, Y_{i'j'}) = \begin{bmatrix} \sigma_{ij}^2 & \sigma_{ii'} \\ \sigma_{ii'} & \sigma_{ij}^2 \end{bmatrix}. \quad (4)$$

Measurements made on different subjects are still assumed to be independent. To fit the model in PROC MIXED, the REPEATED statement is used to specify the repeated measures factor, the subject variable identifying observations that are to be correlated, and a covariance or correlation structure. For the multivariate profile analysis, the unstructured covariance matrix is specified using the TYPE=UN optional parameter.

```
Proc Mixed Data=Vision;
  Where LensStrength="6/06";
  Class Subject Eye;
  Model ResponseTime = Eye;
  Repeated Eye / Subject=Subject Type=UN R RCorr;
  LSMeans Eye / PDiff;
Run;
```

Note that the RANDOM statement is not used in this model. The estimated covariance parameters are given in Table 8. The variance of the difference of the left and right-eye means is again the same as for the previous two models, and the observed test statistic for comparing the left and right eyes is also the same as for the previous analyses.

A correlated errors model can also be fit to the left eye data by removing the RANDOM statement and inserting the appropriate REPEATED statement to the PROC MIXED code given earlier. The four responses measured on a subject under the 4 lens strengths form the multivariate vector of correlated observations.

Table 8. Covariance parameter estimates for the unstructured multivariate correlated errors model.

Cov Parm	Subject	Estimate
UN(1,1)	Subject	12.476
UN(2,1)	Subject	3.905
UN(2,2)	Subject	61.286

```
Proc Mixed Data=Vision;
  Where Eye="Left ";
  Class Subject LensStrength;
  Model ResponseTime = LensStrength / DDFM=KenwardRoger;
  Repeated LensStrength / Subject=Subject Type=UN R RCorr;
  LSMeans LensStrength / PDiff;
Run;
```

The estimated covariance and correlation matrix of the measurements differs somewhat from that obtained using the random effects model. In fact, the unstructured covariance matrix has permitted a negative association ($r=-0.251$) between lens strengths 6/06 and 6/30.

Table 9. Estimated unstructured covariance matrix of errors for subject 1 for the correlated errors model for the left eye data.

Row	Col1	Col2	Col3	Col4
1	12.476	9.262	-7.810	4.952
2	9.262	12.619	10.571	6.524
3	-7.810	10.571	77.476	8.881
4	4.952	6.524	8.881	12.905

Table 10. Estimated unstructured correlation matrix of errors for subject 1 for the correlated errors model for the left eye data.

Row	Col1	Col2	Col3	Col4
1	1.000	0.738	-0.251	0.390
2	0.738	1.000	0.338	0.511
3	-0.251	0.338	1.000	0.281
4	0.390	0.511	0.281	1.000

Now, if the order of lens strengths was randomized for each subject, then we might believe that the measurements should be equi-correlated. Using the correlated errors model, but rather than letting the covariance matrix be unstructured, we might wish to impose an equi-correlated structure on the covariances. This can be done using the TYPE= option of the REPEATED statement. Here, we'll specify compound symmetry (CS) to get the equi-correlated covariance matrix.

```
Proc Mixed Data=Vision;
  Where Eye="Left ";
  Class Subject LensStrength;
  Model ResponseTime = LensStrength / DDFM=KenwardRoger;
  Repeated LensStrength / Subject=Subject Type=CS R RCorr;
  LSMeans LensStrength / PDiff;
Run;
```

An alternative is to fit the heterogeneous variance compound symmetry (CSH) covariance structure.

```
Proc Mixed Data=Vision;
  Where Eye="Left ";
  Class Subject LensStrength;
  Model ResponseTime = LensStrength / DDFM=KenwardRoger;
  Repeated LensStrength / Subject=Subject Type=CSH R RCorr;
  LSMeans LensStrength / PDiff;
Run;
```

MODEL SELECTION

All of these models are potential models for analyzing the data for the left (or right) eye. Because it is the most general of the structures, the multivariate model could be argued as the best (i.e., it makes no structural assumptions). However, the error degrees of freedom for testing for a lens strength effect using this multivariate model is 4 ($F=4.56$, $P>F=0.0883$), while for the homogeneous variance RBD model it is 18 ($F=3.10$, $P>F=0.0529$). Some authors have proposed using information criteria, such as AIC or BIC, for choosing among potential models

(see e.g., Moser and Macchiavelli 2002). Since we are using REML estimation to compare among models, all models in the set must have the same fixed-effects model. This is the case for our lens strength analysis of the left eye data. Table 11 contains the model fit results of these analyses. Here, the model that is selected as "best" using any of the 3 information criteria is the RBD model with heterogeneous variances. This model requires estimation of 5 covariance parameters (half that of the multivariate model), but permits heterogeneous variances and the correlations among pairs of measurements are not required to be equi-correlated.

Table 11. Model fit criteria for 5 random effects/correlated errors models fit to the left eye data.

Model	-2LogL	Parms	AIC	AICc	BIC
Random Subject/Homogeneous Variance	155.5	2	159.5	160.1	159.4
Random Subject/Heterogeneous Variance	141.9	5	151.9	155.2	151.6
Correlated Errors/Unstructured	134.2	10	154.2	171.1	153.7
Correlated Errors/CS	155.5	2	159.5	160.1	159.4
Correlated Errors/CSH	144.4	5	154.4	157.7	154.1

TWO WITHIN-SUBJECT FACTORS

To complete the analysis of this example, we'll consider both of the within-subject factors, eye and lens strength, in a single analysis. A mixed model for two within-subject factors corresponds with a strip-plot design model (see e.g., Schabenberger and Pierce 2002:493-503). The random effects define the three different-sized "experimental units" in this model, allowing covariances to vary due to the subject, eye, and lens strength. This will result in covariances among each of the 8 measurements for a given subject, though different subjects are still assumed to be independent. The PROC MIXED code follows with estimated covariance matrix given in Table 12 and correlation matrix in Table 13.

```
Proc Mixed Data=Vision;
  Class Subject Eye LensStrength;
  Model ResponseTime = Eye | LensStrength / DDFM=KenwardRoger;
  Random Subject Eye(Subject) LensStrength(Subject) / V VCorr;
  LSMeans Eye*LensStrength / PDiff;
Run;
```

Table 12. Estimated covariance matrix of the 8 responses for the strip-plot, two within-subject factors model. Row/Columns 1-4 correspond to left eye lens strength measurements, 5-8 with right eye lens strength measurements.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	48.83	31.97	31.97	31.97	24.31	20.29	20.29	20.29
2	31.97	48.83	31.97	31.97	20.29	24.31	20.29	20.29
3	31.97	31.97	48.83	31.97	20.29	20.29	24.31	20.29
4	31.97	31.97	31.97	48.83	20.29	20.29	20.29	24.31
5	24.31	20.29	20.29	20.29	48.83	31.97	31.97	31.97
6	20.29	24.31	20.29	20.29	31.97	48.83	31.97	31.97
7	20.29	20.29	24.31	20.29	31.97	31.97	48.83	31.97
8	20.29	20.29	20.29	24.31	31.97	31.97	31.97	48.83

Table 13. Estimated correlation matrix of the 8 responses for the strip-plot, two within-subject factors model. Row/Columns 1-4 correspond to left eye lens strength measurements, 5-8 with right eye lens strength measurements.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.000	0.655	0.655	0.655	0.498	0.416	0.416	0.416
2	0.655	1.000	0.655	0.655	0.416	0.498	0.416	0.416
3	0.655	0.655	1.000	0.655	0.416	0.416	0.498	0.416
4	0.655	0.655	0.655	1.000	0.416	0.416	0.416	0.498
5	0.498	0.416	0.416	0.416	1.000	0.655	0.655	0.655
6	0.416	0.498	0.416	0.416	0.655	1.000	0.655	0.655
7	0.416	0.416	0.498	0.416	0.655	0.655	1.000	0.655
8	0.416	0.416	0.416	0.498	0.655	0.655	0.655	1.000

The multivariate, correlated errors model can also be specified using the REPEATED statement with the UN structure. Unfortunately, there are not enough subjects to provide degrees of freedom to estimate the errors. When there are a large number of repeated measurements relative to the number of subjects, the unstructured model may not be feasible. This is one of the reasons that alternative models for repeated measures are needed. As an alternative, PROC MIXED permits direct product structures that are very flexible for two within-subject factors, but do not require as many parameters as the UN structure (for the model below it is 4 versus 36 covariance parameters for the UN multivariate model). By including a random subject effect we can fit a fairly flexible covariance structure.

```
Proc Mixed Data=Vision;
Class Subject Eye LensStrength;
Model ResponseTime = Eye | LensStrength / DDFM=KenwardRoger;
Random Subject / V VCorr;
Repeated Eye LensStrength / Subject=Subject Type=UN@CS R RCorr;
LSMeans Eye*LensStrength / PDiff;
Run;
```

This particular structure is permitting an unstructured matrix (UN) for the levels of eye, and a compound symmetry (CS) structure among the levels of lens strength. The direct product of the UN with the CS structure along with the subject random effect produces the estimated covariance matrix of responses given in Table 14. The corresponding correlation matrix is given in Table 15.

Table 14. Estimated covariance matrix of responses for a subject using the direct product UN@CS covariance structure along with a random subject effect.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	66.290	45.955	45.955	45.955	33.635	30.393	30.393	30.393
2	45.955	66.290	45.955	45.955	30.393	33.635	30.393	30.393
3	45.955	45.955	66.290	45.955	30.393	30.393	33.635	30.393
4	45.955	45.955	45.955	66.290	30.393	30.393	30.393	33.635
5	33.635	30.393	30.393	30.393	50.015	38.199	38.199	38.199
6	30.393	33.635	30.393	30.393	38.199	50.015	38.199	38.199
7	30.393	30.393	33.635	30.393	38.199	38.199	50.015	38.199
8	30.393	30.393	30.393	33.635	38.199	38.199	38.199	50.015

Table 15. Estimated correlation matrix of responses for a subject using the direct product UN@CS covariance structure along with a random subject effect.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.000	0.693	0.693	0.693	0.584	0.528	0.528	0.528
2	0.693	1.000	0.693	0.693	0.528	0.584	0.528	0.528
3	0.693	0.693	1.000	0.693	0.528	0.528	0.584	0.528
4	0.693	0.693	0.693	1.000	0.528	0.528	0.528	0.584
5	0.584	0.528	0.528	0.528	1.000	0.764	0.764	0.764
6	0.528	0.584	0.528	0.528	0.764	1.000	0.764	0.764
7	0.528	0.528	0.584	0.528	0.764	0.764	1.000	0.764
8	0.528	0.528	0.528	0.584	0.764	0.764	0.764	1.000

REPEATED MEASURES IN TIME

When time is the repeated measures factor, then we have additional restrictions with respect to randomization. In the case of the lens powers example, the order that each person was measured at each lens power could have been randomized. Averaging over a large number of subjects in the study, the correlations, after adjusting for any mean differences, among any pair of measurements made at different lens powers should be about equal. However, when time is the factor, we are unable to randomize the order of time itself. There is a strong tendency for measurements on the same subject made near in time to be more highly associated than measurements made further apart in time. Example 2 consists of pig weights measured the same day of each week over a 9-week period (Diggle et al., 1994:35). The ordering of the weeks cannot be randomized so measurements made nearer in time even after adjusting for the growth trend would be expected to be more alike than measurements made further apart in time.

CORRELATED ERRORS

The multivariate unstructured covariance matrix model of correlated errors would be a reasonable model to try with these data. All pigs are measured at the same times and there are enough pigs in the study to permit estimation of the covariance matrix. Unfortunately, we will have to estimate $9 \times 10 / 2 = 45$ covariance parameters. On the other hand, this matrix will give us clues for simpler covariance structure models that we might investigate. We will begin by putting WEEK, the week at which the weight measurement is taken, in the model as a classification variable. This may overfit the trend in growth with time, but should give us a reasonable first look at the covariance matrix with the growth trend adjusted out, since this would be the saturated means model. The PROC MIXED code to fit this model looks like:

```
Proc Mixed Data=Pigs Update;
  Class Week;
  Model Weight = Week / DDFM=KenwardRoger;
  Repeated Week / Subject=Pig Type=UN HLM R RCorr RI;
  LSMeans Week;
Run;
```

The diagonals of Table 16 are the estimated variances of the errors associated with each week's measurements. Note that they consistently increase from 6.10 at week 1 to 40.14 at week 9. This observation is consistent with most growth data, and so covariance structure models that we consider for these data should accommodate heterogeneous variance with time. The correlation matrix of errors, Table 17, should also be studied to determine other potential covariance structures to try.

Table 16. Estimated multivariate unstructured covariance matrix of errors for the saturated means pig growth model.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	6.10	6.31	7.01	7.34	8.39	7.75	8.04	8.38	8.73
2	6.31	7.79	9.02	9.47	11.15	10.37	10.77	10.80	11.74
3	7.01	9.02	12.56	12.68	14.92	14.29	14.87	15.70	17.26
4	7.34	9.47	12.68	13.95	16.29	15.50	16.12	16.80	18.59
5	8.39	11.15	14.92	16.29	20.57	18.60	19.27	19.94	22.57
6	7.75	10.37	14.29	15.50	18.60	19.80	21.32	22.40	25.07
7	8.04	10.77	14.87	16.12	19.27	21.32	24.73	25.86	28.89
8	8.38	10.80	15.70	16.80	19.94	22.40	25.86	29.42	33.32
9	8.73	11.74	17.26	18.59	22.57	25.07	28.89	33.32	40.14

Table 17. Estimated multivariate unstructured correlation matrix of errors for the saturated means pig growth model.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	1.000	0.916	0.802	0.796	0.749	0.705	0.655	0.626	0.558
2	0.916	1.000	0.912	0.908	0.881	0.835	0.776	0.713	0.664
3	0.802	0.912	1.000	0.958	0.928	0.906	0.844	0.817	0.769
4	0.796	0.908	0.958	1.000	0.962	0.933	0.868	0.829	0.786
5	0.749	0.881	0.928	0.962	1.000	0.922	0.855	0.810	0.786
6	0.705	0.835	0.906	0.933	0.922	1.000	0.963	0.928	0.889
7	0.655	0.776	0.844	0.868	0.855	0.963	1.000	0.959	0.917
8	0.626	0.713	0.817	0.829	0.810	0.928	0.959	1.000	0.970
9	0.558	0.664	0.769	0.786	0.786	0.889	0.917	0.970	1.000

The correlation matrix does indicate that the nearer in time that measurements are made, the higher the correlation between them. Thus, we might try the heterogeneous variance autoregressive and antedependence structures for these data. Note that these structures will estimate a separate variance for each week. In order to use the ARH(1) and TOEPH structures, the time intervals between all successive pairs of measurements need to be equal. Methods to deal with unequally spaced intervals will be considered later.

Table 18. Covariance parameter estimates for the heterogeneous variance first-order autoregressive structure for the saturated means model.

Cov Parm	Subject	Estimate
Var(1)	Pig	7.6605
Var(2)	Pig	9.7202
Var(3)	Pig	14.3295
Var(4)	Pig	14.4184
Var(5)	Pig	20.5870
Var(6)	Pig	18.7938
Var(7)	Pig	21.7207
Var(8)	Pig	24.7070
Var(9)	Pig	33.4050
ARH(1)	Pig	0.9468

Table 19. Covariance parameter estimates for the heterogeneous variance general-order toeplitz structure for the saturated means model.

Cov Parm	Subject	Estimate
Var(1)	Pig	7.9282
Var(2)	Pig	10.0145
Var(3)	Pig	14.4798
Var(4)	Pig	14.4616
Var(5)	Pig	20.5005
Var(6)	Pig	18.6084
Var(7)	Pig	21.2504
Var(8)	Pig	24.1216
Var(9)	Pig	32.9290
TOEPH(1)	Pig	0.9465
TOEPH(2)	Pig	0.8947
TOEPH(3)	Pig	0.8409
TOEPH(4)	Pig	0.7844
TOEPH(5)	Pig	0.7271
TOEPH(6)	Pig	0.6568
TOEPH(7)	Pig	0.6059
TOEPH(8)	Pig	0.5287

Table 20. Covariance parameter estimates for the first-order antedependence structure for the saturated means model.

Parm	Subject	Estimate
Var(1)	Pig	6.0953
Var(2)	Pig	7.7862
Var(3)	Pig	12.5611
Var(4)	Pig	13.9464
Var(5)	Pig	20.5655
Var(6)	Pig	19.8004
Var(7)	Pig	24.7323
Var(8)	Pig	29.4228
Var(9)	Pig	40.1373
Rho(1)	Pig	0.9156
Rho(2)	Pig	0.9118
Rho(3)	Pig	0.9582
Rho(4)	Pig	0.9621
Rho(5)	Pig	0.9219
Rho(6)	Pig	0.9633
Rho(7)	Pig	0.9586
Rho(8)	Pig	0.9695

The PROC MIXED code to estimate the heterogeneous variance first-order autoregressive structure for the saturated means model is:

```
Proc Mixed Data=Pigs Update;
  Class Week;
  Model Weight = Week / DDFM=KenwardRoger;
  Repeated Week / Subject=Pig Type=ARH(1) R RCorr RI;
  LSMeans Week;
Run;
```

The estimated variances for each of the weeks given in Table 18 for the heterogeneous variance first-order autocorrelation structure are indicated by Var(*), while ARH(1) is the correlation between pairs of measurements made on the same pig one time interval apart. If you want to compute the correlation for a different interval of time, raise the correlation to the power of the desired interval. For example, the correlation between measurements made 2 weeks apart would be $0.9468^2=0.8964$.

The heterogeneous toeplitz structure permits a general-order autoregressive model to be fit. Not only will there be separate variance estimates for each week, but there will be a correlation parameter estimated for each time interval that measurements can be separated by (Table 19). This structure is specified by replacing the TYPE=ARH(1) with TYPE=TOEPH in the REPEATED statement of the saturated means model. The first-order antedependence structure is specified by TYPE=ANTE(1). Like the ARH(1) structure, ANTE(1) models first-order dependence, but permits that dependence to depend upon which pair of adjacent weeks are being referenced. These first-order structures are also models for conditional independence because any pair of measurements made on the same subject that are further apart than one time interval are conditionally independent given the intervening time measurements (Macchiavelli and Moser 1997). This can be more readily observed by noticing the zeros present in the concentration matrix, or inverse of the covariance matrix (Table 21). Thus, we might think also about studying the concentration matrix for the unstructured covariance model (Table 22). It tends to have larger (absolute) values along the first minor diagonal band as we expect for a first-order model, but there are some other elements in the matrix that suggests that some higher-order dependence is present.

Table 21. Estimated concentration matrix or inverse of the covariance matrix for the first-order antedependence structure for the saturated means model. An empty cell indicates that the value is zero.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	1.015	-0.822							
2	-0.822	1.428	-0.547						
3		-0.547	1.365	-0.885					
4			-0.885	1.768	-0.763				
5				-0.763	0.929	-0.305			
6					-0.305	0.987	-0.604		
7						-0.604	1.019	-0.438	
8							-0.438	0.950	-0.469
9								-0.469	0.414

Table 22. Estimated concentration matrix for the unstructured covariance model for the saturated means model.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	1.34	-1.42	0.18	-0.14	0.09	0.19	0.26	-0.55	0.21
2	-1.42	2.44	-0.61	-0.14	-0.14	-0.11	-0.48	0.74	-0.20
3	0.18	-0.61	1.24	-0.65	-0.05	-0.09	0.25	-0.35	0.11
4	-0.14	-0.14	-0.65	1.96	-0.65	-0.44	0.10	-0.11	0.10
5	0.09	-0.14	-0.05	-0.65	0.79	-0.26	0.02	0.25	-0.16
6	0.19	-0.11	-0.09	-0.44	-0.26	1.64	-0.75	-0.15	0.02
7	0.26	-0.48	0.25	0.10	0.02	-0.75	1.13	-0.70	0.15
8	-0.55	0.74	-0.35	-0.11	0.25	-0.15	-0.70	1.53	-0.72
9	0.21	-0.20	0.11	0.10	-0.16	0.02	0.15	-0.72	0.51

Again, the information criteria could be used to help select a covariance structure to use in modeling these data. Table 23 shows that the antedependence model has the best fit of these 4 models based upon the AIC and AICC criteria, while the ARH(1) structure fits best if you use the BIC criterion. Once a covariance structure is selected, then models for the trend can be explored in greater detail.

Table 23. Model fits of the covariance structures using the saturated means model.

Structure	-2RLogL	AIC	AICC	BIC
UN	1505.6	1595.6	1606.6	1679.8
ARH(1)	1576.2	1596.2	1596.7	1614.9
TOEPH	1567.2	1601.2	1602.7	1633.0
ANTE(1)	1553.1	1587.1	1588.7	1619.0

Consider the model where we now use a first-order antedependence covariance structure, but model mean pig weight as a straight-line function of time. We can specify this model for PROC MIXED as:

```
Proc Sort Data=Pigs;
  By Pig Week;
Run;

Proc Mixed Data=Pigs Update;
  Model Weight = Week / DDFM=KenwardRoger;
  Repeated / Subject=Pig Type=ANTE(1) R RCorr;
Run;
```

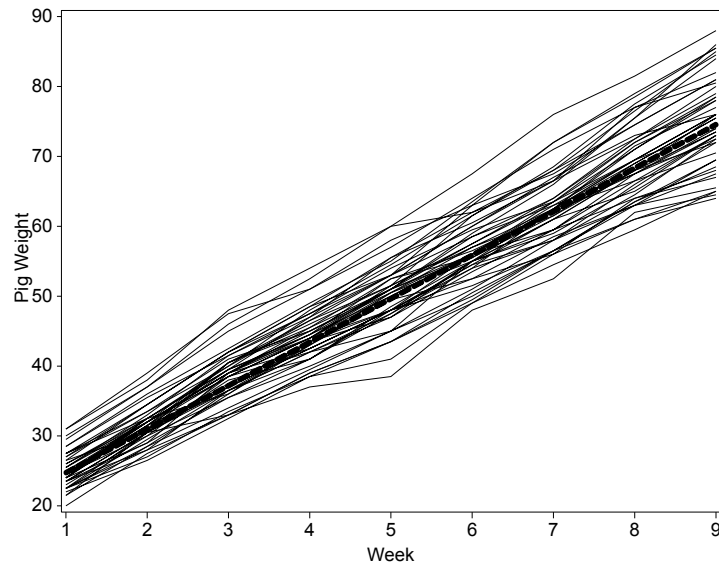
To insure that measurements are correctly matched to the appropriate week, say if you had missing values, you could alternatively specify the model using week as both categorical and continuous by creating a copy of the week variable:

```
Data Pigs;
  Set Pigs;
  Week1=Week;
Run;

Proc Mixed Data=Pigs Update;
  Class Week;
  Model Weight = Week1 / DDFM=KenwardRoger;
  Repeated Week / Subject=Pig Type=ANTE(1) R RCorr;
Run;
```

A plot of the observed and predicted values (Figure 1) indicates a high degree of individual variability about the predicted line. A technique to fit subject-specific growth models might be useful, especially for making predictions about pigs currently in the study. Random coefficient regression models can address this need.

Figure 1. Observed pig weight growth profiles with fitted linear growth model (dotted line).



RANDOM COEFFICIENTS

A random coefficient regression model can be used to fit subject-specific models about a population-averaged model. Consider a simple linear regression model to model the weekly pig weights, but permit both the intercept and slope to have a bivariate normal distribution with marginal means describing the population-average growth model. The model can be written as

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \varepsilon_{ij} \quad (5)$$

where $E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}$ is the linear population-averaged growth model, and b_{0i} and b_{1i} are random effects with zero mean and covariance matrix

$$\text{Cov}(b_{0i}, b_{1i}) = \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_0 b_1} \\ \sigma_{b_0 b_1} & \sigma_{b_1}^2 \end{bmatrix}. \quad (6)$$

The population-averaged parameters β_0 and β_1 , and the parameters of the random effects covariance matrix along with the residual variance are to be estimated. This can be accomplished using PROC MIXED via the code:

```
Proc Sort Data=Pigs;
  By Pig Week;
Run;

Proc Mixed Data=Pigs Update;
  Model Weight = Week / Solution OutPM=PAFit OutP=SSFit;
  Random Intercept Week / Solution Subject=Pig Type=UN G GCorr V VCorr;
  Estimate "Week 1 Ave Weight" Intercept 1 Week 1;
  Estimate "Week 9 Ave Weight" Intercept 1 Week 9;
Run;
```

The SOLUTION option in the MODEL statement will list the population-averaged regression coefficients, described in the output as the fixed effects solution (Table 24), while the covariance parameter estimates section and the G option of the RANDOM statement will list the variance components (Table 25). The OUTPM= and OUTP= data set options on the MODEL statement produce data sets of the predicted means (population-averaged model) and individual predicted values (subject-specific model). The ESTIMATE statements can be used to predict the population mean value for specific times. The subject-specific and population-averaged fits are plotted in Figure 2. Compare these fits with the observed growth profiles given in Figure 1.

Figure 2. Subject-specific and population-averaged model fits for the random coefficient regression model of pig weights.

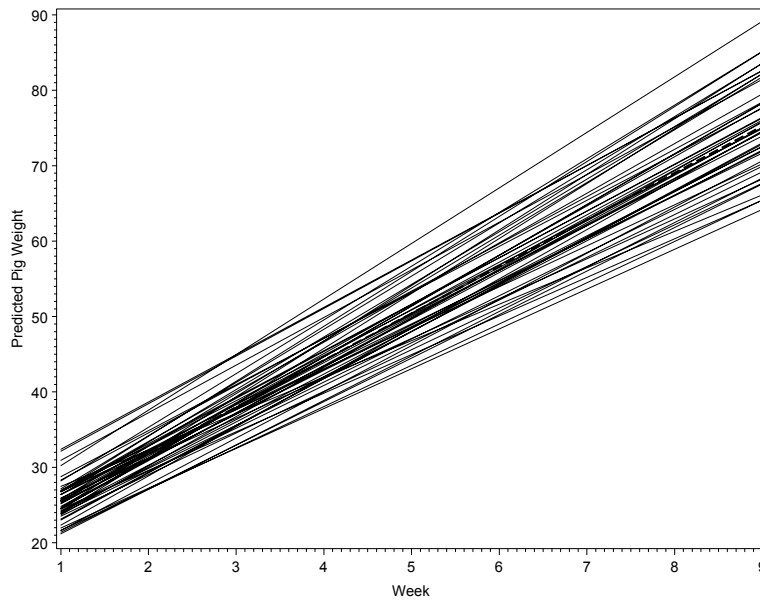


Table 24. Solution of the fixed effects for the random coefficient regression model.

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	19.3556	0.4039	47	47.93	<.0001
Week	6.2099	0.09204	47	67.47	<.0001

Table 25. Estimated G matrix or covariance matrix of the random effects for the random coefficient regression model.

Row	Effect	Subject	Col1	Col2
1	Intercept	1	6.9865	-0.1034
2	Week	1	-0.1034	0.3800

When the estimated covariance matrix of weight measurements is computed (Table 26), the matrix is observed to have properties that we desire, i.e., the variance of weights increases with time, and when the correlations are computed they show that observations measured near in time are more highly correlated than those measured further apart in time. Thus, we have a fairly simple model that captures some of the qualities of the variability of this data.

The SOLUTION option of the RANDOM statement produces a table of the EBLUPs (empirical best linear unbiased predictors) associated with the random effects, and these correspond with the b_{0i} and b_{1i} in the random coefficient regression model. A portion of this output is displayed in Table 27. The EBLUPs, together with the solution for the fixed effects, can be used to compute the subject-specific equations. Note that this technique has a lot of applicability to repeated measures problems where subjects are measured at different, irregular, and unequally-spaced intervals of time. Also, random coefficients can be easily extended to nonlinear models.

Table 26. Estimated covariance matrix of weights of pigs for the 9 weeks of measurement based upon the random coefficient regression model.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	8.76	7.44	7.71	7.99	8.27	8.54	8.82	9.10	9.37
2	7.44	9.69	8.75	9.41	10.06	10.72	11.38	12.03	12.69
3	7.71	8.75	11.38	10.82	11.86	12.90	13.93	14.97	16.01
4	7.99	9.41	10.82	13.84	13.66	15.07	16.49	17.91	19.32
5	8.27	10.06	11.86	13.66	17.05	17.25	19.05	20.84	22.64
6	8.54	10.72	12.90	15.07	17.25	21.02	21.60	23.78	25.96
7	8.82	11.38	13.93	16.49	19.05	21.60	25.76	26.72	29.27
8	9.10	12.03	14.97	17.91	20.84	23.78	26.72	31.25	32.59
9	9.37	12.69	16.01	19.32	22.64	25.96	29.27	32.59	37.50

Table 27. Solution to the random effects or empirical best linear unbiased predictors for the random coefficient regression model for the first 3 pigs.

Effect	Subject	Estimate	Std Err Pred	DF	t Value	Pr > t
Intercept	1	0.2402	0.9239	336	0.26	0.7950
Week	1	-0.3964	0.1747	336	-2.27	0.0239
Intercept	2	-1.5915	0.9239	336	-1.72	0.0859
Week	2	0.5114	0.1747	336	2.93	0.0036
Intercept	3	-3.5375	0.9239	336	-3.83	0.0002
Week	3	0.3218	0.1747	336	1.84	0.0663

CORRELATED ERRORS – CONTINUOUS TIME/CONTINUOUS SPACE

Consider now that the pig weights are measured irregularly for each pig, such that the time intervals between measurements would be more or less unique to each pig. Under these circumstances we will need a continuous-time model to describe the covariances among the errors. The spatial data covariance structures provided in PROC MIXED can be used for this purpose. The distance used in these functions will be the continuous time between measurements. As an example the spatial power covariance function has the form

$$\text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j}) = \sigma_\varepsilon^2 \rho^{d_{ii'}} \quad (7)$$

where $d_{ii'}$ is the distance in time between weeks i and i' , and ρ is a correlation parameter that measures the linear association between the errors measured 1 unit of distance (here, time) apart. Note that when the time intervals are integers, this model reduces to the first-order autoregressive model AR(1). The PROC MIXED code to fit this structure to the pigs data follows. In addition to correlated errors, a random slope coefficient will also be included to permit the variance of the weights to vary with time.

```
Proc Sort Data=Pigs;
  By Pig Week;
Run;

Proc Mixed Data=Pigs Update;
  Model Weight = Week / Solution OutPM=PAFit OutP=SSFit;
  Random Week / Solution Subject=Pig V VCorr;
  Repeated / Subject=Pig Type=SP(POW) (Week) R=1 RCorr=1;
Run;
```

Note the optional “=1” added to both the R and RCORR options on the REPEATED statement. If the subjects are truly measured at irregular intervals and have unequal numbers of measurements per subject, then the covariance matrix reported will depend upon the subject. Thus you can specify the particular subject or subjects that you would like listed. Since the pigs are measured equally and at the same times, any one of the subjects in this example will provide the covariance matrix of interest. The estimate of ρ is 0.8385 (Table 28) and corresponds with the first-order autocorrelation parameter. The estimated covariance matrix for the weights does indeed have a non-constant variance with time (Table 29), and so is consistent with the observed data in this regard.

In Table 30 the model fits of the first-order antedependence correlated errors model (ANTE(1)), random intercept and slope coefficient regression model (RC), and random slope plus spatial power correlated errors model (POW+RC) are compared. These models all assumed a straight-line (linear) model for the trend. Using the AIC criteria the ANTE(1) model is selected as best of the 3, while the BIC criterion selects the POW+RC model as best of the 3.

If the repeated measurements are repeated in space, then the spatial covariance structures can be directly applied. If there are many spatial measurements per subject, then spatial data techniques should be considered (see e.g., Schabenberger and Pierce 2002:561-625), including the use of PROC VARIOGRAM. A PARMs statement should also be used so as to provide a search region that can help insure that the final iteration is at the maximum of the likelihood. For the pigs example there are 3 parameters (Table 28) for which values will need to be supplied if the PARMs statement is used. The order of the parameters matches the covariance parameter estimates table as in Table 28.

```
Proc Mixed Data=Pigs Update;
  Model Weight = Week / Solution OutPM=PAFit OutP=SSFit;
  Random Week / Solution Subject=Pig V VCorr;
  Repeated / Subject=Pig Type=SP(POW) (Week) R=1 RCorr=1;
  Parms (0.2 To 0.8 By 0.1) (0.5 To 0.9 By 0.1) (4 To 8 By 2);
Run;
```

For the power spatial structure, starting parameter values are not difficult as ρ is a correlation. For other spatial structures, however, a working knowledge of spatial variograms is very useful (Moser and Macchiavelli 1996, Schabenberger and Pierce 2002).

Table 28. Covariance parameter estimates corresponding to the spatial power covariance structure coupled with a random slope coefficient model.

Cov Parm	Subject	Estimate
Week	Pig	0.3930
SP(POW)	Pig	0.8385
Residual		6.3258

Table 29. Estimated covariance matrix of weight measurements corresponding to the spatial power covariance structure coupled with a random slope coefficient model.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	6.72	6.09	5.63	5.30	5.09	4.98	4.95	4.99	5.08
2	6.09	7.90	7.66	7.59	7.66	7.84	8.12	8.49	8.92
3	5.63	7.66	9.86	10.02	10.34	10.80	11.38	12.05	12.81
4	5.30	7.59	10.02	12.61	13.16	13.88	14.73	15.70	16.77
5	5.09	7.66	10.34	13.16	16.15	17.09	18.20	19.45	20.81
6	4.98	7.84	10.80	13.88	17.09	20.47	21.81	23.31	24.95
7	4.95	8.12	11.38	14.73	18.20	21.81	25.58	27.31	29.21
8	4.99	8.49	12.05	15.70	19.45	23.31	27.31	31.48	33.60
9	5.08	8.92	12.81	16.77	20.81	24.95	29.21	33.60	38.16

Table 30. Model fits for 3 linear growth models fitted to the pig weight data. ANTE(1) is a first-order antedependence correlated errors model, RC is a random intercept and slope coefficient model, and POW+RC is a random slope with a power spatial covariance correlated errors model.

Structure	-2RLogL	AIC	AICC	BIC
ANTE(1)	1603.6	1637.6	1639.1	1669.4
RC	1740.9	1748.9	1749.0	1756.4
POW+RC	1645.5	1651.5	1651.5	1657.1

EXTENSIONS TO GENERALIZED LINEAR MODELS

Diggle et al. (1994:13-14) present data from a study on 59 epileptics, where the number of seizures is recorded over a baseline period, then subjects are randomized to either receive progabide, an anti-epileptic drug, or a placebo. The number of seizures is then recorded for each subject over 4 consecutive two-week periods. The age of each subject is also included in the data set. The response variable is a non-negative discrete (integer) random variable. If we had counts for only a single time period, then we might consider a generalized linear model analysis of these data (e.g., PROC GENMOD). With repeated measurements, we could use the generalized estimating equations approach provided through the REPEATED statement of PROC GENMOD, or we could use the generalized linear mixed model approach provided through the %GLIMMIX macro (Littell et al. 1996:423-460, Kuss 2002). This macro, available in the SAS/STAT sample library in the SAS installation directory tree, or at the SAS support web site (<http://ftp.sas.com/techsup/download/stat/glmm800.html>), permits fitting of generalized linear models but with the added features of PROC MIXED, specifically the RANDOM and REPEATED statements. The %GLIMMIX macro uses a quasi-likelihood approach that iteratively passes pseudo-data as adjusted by the quasi-likelihood approach to the MIXED procedure (Littell et al. 1996:423-460, Kuss 2002).

Consider a Poisson generalized linear model for the epileptic fits data where a log link function relates the Poisson mean to a linear function of the baseline measurement, patient's age, treatment, period of measurement, and the interaction of the treatment with period. Since there are 4 measurements made on each person, include a random subject effect in the model and specify a covariance structure on the errors. The random effect will not only permit association among the measurements, but will account for some of the overdispersion in the data. By default, %GLIMMIX will include a single parameter for overdispersion.

The code for one possible overdispersed Poisson mixed model with correlated errors is:

```
%Include "glimmix.sas";

%Glimmix(Data=EpilepticFits,
  Stmts=%Str(
    Class Treatment Period Subject;
    Model NumberOfFits = Age BaseLine Treatment Period Treatment*Period
      / DDFM=KenwardRoger;
    Random Subject(Treatment);
    Repeated Period / Subject=Subject(Treatment) Type=AR(1);
  ),
  Error=Poisson,
  Link=Log,
  Out=PoissonResults
)
Run;
```

Note that the program code given as the argument to the STMTS= parameter matches code contained within a PROC MIXED call. In addition the distribution and link function are specified. Omission of both the RANDOM and REPEATED statements in the above code would fit a model that ignores the dependence among the repeated observations, and would lead to an inference of a highly significant treatment effect ($P < 0.0001$) and significant period effect ($P = 0.0192$). When the code above is used to model the dependence among measurements made on the same subject, neither the treatment ($P = 0.1224$) nor period ($P = 0.1882$) effects are significant. These latter values better account for the variation in the data, and adjust the degrees of freedom for the number of subjects in the study. Note that the random coefficient and continuous-time covariance structure models can work with the %GLIMMIX macro as well. Diggle et al. (1994) fit alternative models to these data that you might try.

EXTENSIONS TO NONLINEAR MODELS

Diggle et al. (1994:5-7) report data on the size growth of Sitka spruce trees grown under either an ozone enriched environment or a control environment. Trees are measured at irregularly spaced intervals, and tree size is constructed as the product of diameter squared with height. Trees are grown in growth chambers with 2 chambers assigned to each treatment. Trees are followed for 2 growing seasons. In the following example growth data for the control group of 25 trees, up to the beginning of the second growth phase, will be used. The raw growth profiles are shown in Figure 3 and suggest that an asymptotic growth model might describe growth through the first phase. The von Bertalanffy or Mitscherlich growth model (Schabenberger and Pierce 2002:229) could be used to model the size growth through this phase. It has the form

$$Y_{ij} = A \left(1 - \exp(-k(t_{ij} - t_0)) \right) + \varepsilon_{ij} \quad (8)$$

where Y_{ij} is the size of the i^{th} tree at the j^{th} time, t_{ij} is the number of days of growth of the i^{th} tree at the j^{th} time, A is the asymptotic size parameter, k is a growth rate controlling parameter, and t_0 is an intercept adjustment parameter for the time when size=0. Note that the asymptotes for the individual trees vary considerably, so a non-linear, random coefficient regression model with random asymptotes might be a good first model. In addition, a correlated errors model will be considered so as to account for additional variation due to repeated measurements.

The %NLINMIX macro permits fitting of non-linear mixed models and makes use of the features in PROC MIXED. This technology is described in Littell et al. (1996:461-489), and the macro is available in the SAS/STAT sample library in the SAS installation directory tree, or at the SAS support web site (<http://ftp.sas.com/techsup/download/stat/nlmm800.html>). To incorporate the random asymptote in the growth model, the asymptote parameter is rewritten to have a population-averaged parameter and a subject-specific parameter,

$$Y_{ij} = (A + b_i) \left(1 - \exp(-k(t_{ij} - t_0)) \right) + \varepsilon_{ij} \quad (9)$$

where b_i will be the subject-specific amount by which the i^{th} tree's asymptote deviates from the population averaged asymptote. The macro actually uses pseudo-data constructed from the original data that it passes to PROC MIXED in an iterative process to estimate the parameters. Thus the covariance structures are not placed directly on the errors of equation (9), but on the errors of the pseudo-data.

The code for fitting the above random coefficient model with a spatial power correlated errors structure is:

```
%NLINMIX(Data=Season1,
  Response=Size,
  Subject=Tree,
  Model=%Str(
    expr1=exp(-k*(Time-t0)/400);
    expr2=1-expr1;
    Pred = (A+b1)*(expr2);
  ),
  Derivs=%Str(
    D_A = expr2;
    D_k = (A+b1)*(Time-t0)/400*expr1;
    D_t0 = -(A+b1)*k/400*expr1;
    D_b1 = D_A;
  ),
  ParmS=%str(
    A =1000
    k =3
    t0=100),
  Random=b1,
  Expand=EBLUP,
  RType=SP(POW)(Time),
  RSub=Tree
)
Run;
```

Unlike the %GLIMMIX macro, the %NLINMIX macro has statements that match PROC NLIN syntax, but has parameters such as RANDOM=, RTYPE=, and RSUB= that are used with the calls to PROC MIXED. The solution to the fixed effects (Table 31) gives the parameter estimates for the population-averaged model, while the solution to the random effects (Table 32) gives the EBLUPs that can be used to obtain the subject-specific estimated growth curves that are plotted in Figure 4.

Table 31. Solution for the fixed effects for the von Bertalanffy growth curve model fitted to the Sitka spruce trees data. Note that derivative names of the parameters are given in the table.

Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
D_A	430.08	46.4128	148	9.27	<.0001	0.05	338.36	521.8
D_K	4.1195	0.2389	148	17.24	<.0001	0.05	3.6474	4.5916
D_T0	131.19	2.2311	148	58.8	<.0001	0.05	126.78	135.6

Table 32. Solution for the random effects for the first 5 trees for the von Bertalanffy growth curve model fitted to the Sitka spruce trees data. Note that the derivative name of the parameter is given in the table.

Effect	_subject	Estimate	Std Err Pred	DF	t Value	Pr > t
D_B1	1	-110.9	52.064	148	-2.13	0.0348
D_B1	2	348.4	52.223	148	6.67	<.0001
D_B1	3	-145.7	52.077	148	-2.80	0.0058
D_B1	4	54.0	52.050	148	1.04	0.3016
D_B1	5	-6.2	52.046	148	-0.12	0.9050

Figure 3. Raw growth profiles of the Sitka spruce trees grown under the control conditions through the first growth phase.

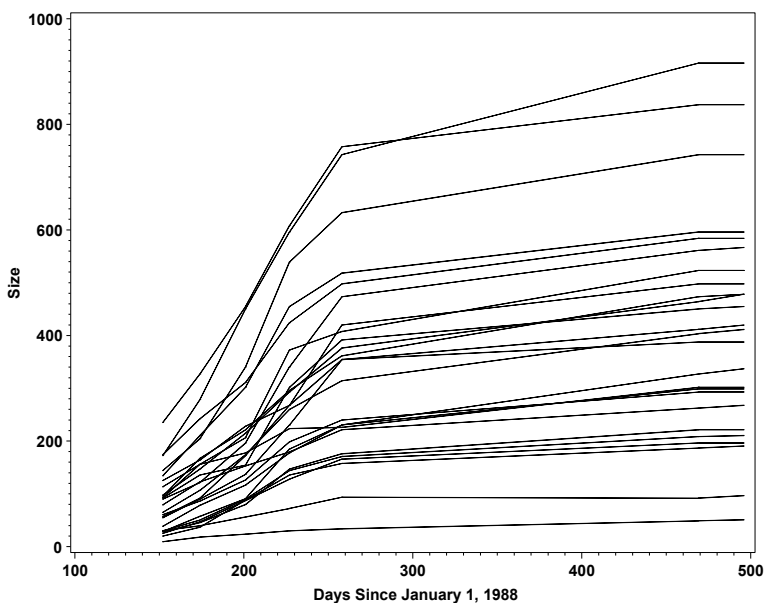
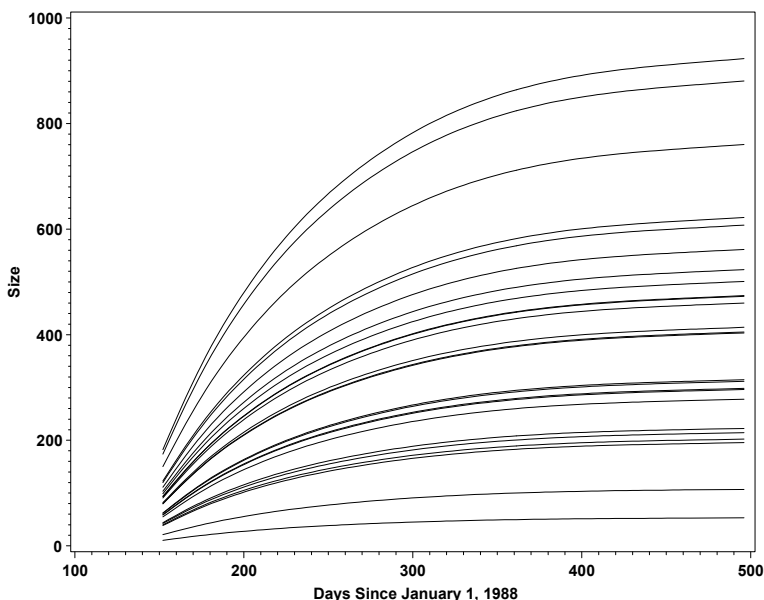


Figure 4. Predicted growth profiles of the Sitka spruce trees grown under the control conditions through the first growth phase using a random coefficient asymptote and a spatial power correlated errors model.



CONCLUSION

As a part of my consulting mission at the University I am working with a graduate student who is studying root growth of a marsh plant grown under varying experimental conditions in rhizotrons, a device that permits direct, nondestructive observation of root growth. Periodically, but not at equally spaced intervals of time, roots are counted at 4 different depths below the surface of soil in each rhizotron. Further each counted root is characterized into 1 of 3 size classes. Thus there are repeated measurements made in both time and space on each rhizotron, the response is a discrete random variable, the roots may transition from one size class to another through time, and the number of rhizotrons is small compared with the large number of measurements. Further, the plants are growing, and so the trend in roots counts may follow a non-linear trajectory. Flexible repeated measures modeling approaches available through software such as PROC MIXED are essential for capturing the many facets to experiments such as this one.

This paper has not dealt with the many diagnostic analyses that should be incorporated into the use of this repeated measures technology. There are many graphical aids to explore repeated measures data. Residuals and EBLUPs could be examined for appropriateness of the normal distribution model. The effects of influential observations and outliers on inference should also be considered. For non-linear models, including generalized linear mixed models, the NL MIXED procedure should also be considered as an alternative to the %GLIMMIX and %NLINMIX macros for repeated measurements data. Users that want to analyze repeated measures with PROC MIXED, should consult the references that follow as these will describe many of the graphical and diagnostic aids omitted in this paper, along with the necessary background information and alternative modeling strategies. Additional SAS code for repeated measures is given in Kuss (2002), Littell et al. (1996), Schabenberger and Pierce (2002), and Verbeke and Molenberghs (2000).

REFERENCES

- Diggle, P. J., Liang, K., and Zeger, S. L. (1994), *Analysis of Longitudinal Data*, Oxford: Clarendon Press, 253pp.
- Johnson, R. A. and Wichern, D. W. (2002), *Applied Multivariate Statistical Analysis*, Fifth Edition, Upper Saddle River, NJ : Prentice Hall, 767pp.
- Kuss, O. (2002), "How to Use SAS® for Logistic Regression with Correlated Data," *Proceedings of the Twenty-seventh Annual SAS Users Group International Conference*, paper 261-27, April 2002, <<http://www2.sas.com/proceedings/sugi27/p261-27.pdf>> (January 30, 2004).
- Lindsey, J.K. (1993), *Models for Repeated Measurements*, Oxford: Clarendon Press, 413pp.
- Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996), *SAS® System for Mixed Models*, Cary, NC: SAS Institute, Inc., 633pp.
- Macchiavelli, R. E. and Moser, E. B. (1997), "Analysis of repeated measures with antedependence covariance models," *Biometrical Journal* 39 (3), 339-350.
- Moser, E. B. and Macchiavelli, R. E. (1996), "Methods for Spatial Analysis," Chapter 5 (pages 91-113) in *Geographical Information Systems in Hydrology*, Singh, V. P. and Fiorentino, M., Eds., Water Science and Technology Library Volume 26, The Netherlands: Kluwer Academic Publishers.
- Moser, E. B., and Macchiavelli, R. E. (2002), "Model selection techniques for repeated measures covariance structures," *Proceedings of Fourteenth Annual Kansas State University Conference on Applied Statistics in Agriculture*, 17-31.
- Schabenberger, O. and Pierce, F. J. (2002), *Contemporary Statistical Models for the Plant and Soil Sciences*, Boca Raton, FL: CRC Press, 738pp.
- Verbeke, G. and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer-Verlag, 568pp.

CONTACT INFORMATION

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

Edgar Barry Moser
 Louisiana State University
 Department of Experimental Statistics
 Baton Rouge, LA 70803-5606
 Work Phone: 225-578-8303
 Fax: 225-578-8344
 Email: bmoser@lsu.edu
 Web: www.stat.lsu.edu/faculty/moser/

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.