Modeling Variation in Repeated Measures Data
Ramon C. Littell, Department of Statistics, University of Florida

ABSTRACT
Statistical analysis of repeated measures data has been a problem because of covariation among measurements on the same experimental unit. Until recently, analysis techniques available in computer software only permitted the user to ignore or avoid covariance structure rather than incorporate it into the statistical model. Ignoring covariance structure results in erroneous inference and avoiding it results in suboptimal inference. The MIXED procedure provides a rich selection of modeling structures through the RANDOM and REPEATED statements. Examples will illustrate how to choose a covariance structure, and the effects on significance tests and confidence intervals.

INTRODUCTION
Mixed linear statistical models state that observed data consist of two parts, fixed effects and random effects. Fixed effects define the mean of the population from which the observation was drawn, and the random effects define the variance of the observation and its covariance with other observations. Mixed linear models are often used with repeated measures data to accommodate the covariation between observations on the same subject at different times. PROC MIXED in the SAS System provides a rich selection of covariance structure from which to choose. Most discussion in the literature has focused on the importance of proper covariance modeling in order to obtain valid F tests for testing time effects and treatment-by-time interaction effects. In the present paper we examine the effects of choice of covariance structure on estimates of treatment means at various times, and on standard errors of differences between treatment means.

EXAMPLE DATA SET
An example from the pharmaceutical industry, which compared the effects of two drugs and a placebo on a measure of respiratory ability, called FEV1, will be used to illustrate covariance modeling. Twenty-four patients were assigned to each of the treatment groups, and FEV1 was measured at baseline (immediately prior to administration of the drugs), and at hourly intervals thereafter for eight hours. Data were analyzed using PROC GLM and PROC MIXED. Two SAS data sets were created. The first data set, named FEV1MULT, contained data in a multivariate mode and had variables DRUG, PATIENT, BASEFEV1, and FEV11H through FEV18H. The second data set, named FEV1UNI, contained data in a univariate mode with variables DRUG, PATIENT, HR, BASEFEV1, and FEV1. Treatment means are plotted versus HR in Figure 1. The graph shows that means for the three treatment groups are essentially the same at HR=0 (baseline). At HR=1 the mean for drug B is larger than the mean for drug A, and both of the drug means are much larger than the placebo mean. Means for drugs A and B continue to be larger than the placebo means for subsequent hours, but the magnitudes of the differences decrease sharply with time. It is of interest to estimate differences between the treatment group means at various times, and to estimate differences between means for the same treatment at different times.

COMPARISON OF COVARIANCE STRUCTURES
We first use PROC GLM with a REPEATED statement to examine the covariance structure in the data. The variables FEV1-FEV6 are regarded as repeated measures and the variable BASEFEV1 is considered a baseline covariable. Run the statements

```
proc glm data=fev1mult; class drug;
model fev1_1h-fev1_8h = fev0_0 drug;
```
contrast 'trt vs cont' drug 1 1 -2;
contrast 'drga vs drgb' drug 1 -1 0;
repeated time contrast / summary printe;
run;

The within-subject correlation matrix is printed because of the PRINTE option, and is shown in Table 1. The correlations between FEV1_1 and FEV1_2-FEV1_8 are in the first row of the matrix. Correlations generally decrease from 0.893 with FEV1_2 down to 0.642 with FEV1_8. Similar decreases are found between FEV1_2 and FEV1_3-FEV1_8, between FEV1_3 and FEV1_4-FEV1_8, etc. In short, correlations between pairs of FEV1 measurements decrease with the number of hours between the times the measurements were obtained. This is a common phenomenon with repeated measures data. As a consequence, a univariate analysis of variance is likely not appropriate. A formal test of whether the covariance structure meets the necessary assumptions (called the Huyn-Feldt conditions) for univariate ANOVA is given by the test for sphericity applied to orthogonal components. The approximate chi-square form of the test has the value 183.98 with 27 degrees of freedom, and has p-value equal to 0.0001. According to this test, the conditions are not met. Thus another type of analysis must be used.

<table>
<thead>
<tr>
<th>Table 1. Partial Correlations of FEV1 at 8 Times from REPEATED Statement in PROC GLM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>

We now turn to PROC MIXED for analyses which accommodate structures defined on the covariance matrix. Define Y_{ijk} to be FEV1 measured at HR = k on PATIENT = j in DRUG = i. The structures in discussion pertain to covariances among measures at different hours on the same patient. Measures on different patients are considered independent in all cases. We consider five covariance structures, defined as follows:

1. simple \( \text{Cov}(Y_{ijk}, Y_{ijl}) = 0 \text{ if } k \neq l \)
   \( = \sigma^2 \text{ if } k = l \)
2. compound symmetric \( \text{Cov}(Y_{ijk}, Y_{ijl}) = \sigma^2 \text{ if } k \neq l \)
   \( = \sigma_1^2 + \sigma_2^2 \text{ if } k = l \)
3. autoregressive \( \text{Cov}(Y_{ijk}, Y_{ijl}) = \sigma^2 p^{|k-l|} \)
4. autoregressive with random effect for patient \( \text{Cov}(Y_{ijk}, Y_{ijl}) = \sigma_1^2 + \sigma_2^2 p^{|k-l|} \)
5. unstructured \( \text{Cov}(Y_{ijk}, Y_{ijl}) = \sigma_{kl} \)

Models with these covariance structures are implemented with the respective sets of statements:

1. proc mixed data=fev1uni; class drug patient time;
   model fev1 = fev1_0 drug time drug'time;
   repeated / type = simple sub = patient r rcorr;
run;
2. proc mixed data=fev1uni; class drug patient time;
   model fev1 = fev1_0 drug time drug'time;
   repeated / type = cs sub = patient r rcorr;
run;
3. proc mixed data=fev1uni; class drug patient time;
   model fev1 = fev1_0 drug time drug'time;
   repeated / type = ar(1) sub = patient r rcorr;
run;
4. proc mixed data=fev1uni; class drug patient time;
   model fev1 = fev1_0 drug time drug'time;
   random patient;
   repeated / type = ar(1) sub = patient r rcorr;
run;
5. proc mixed data=fev1uni; class drug patient time;
   model fev1 = fev1_0 drug time drug'time;
   repeated / type = un sub = patient r rcorr;
run;

Rather than display results for these statements in sequence, we display comparable parts in groups. First, parameter estimates in the covariance matrices for the various structures (excepting "unstructured") are:

1. simple \( \sigma^2 = 0.267 \)
2. compound symmetric \( \sigma^2 = 0.206 \)
   \( \sigma_2 = 0.063 \)
3. autoregressive \( \sigma^2 = 0.266 \)
   \( \rho = 0.856 \)
4. autoregressive with random effect for patients \( \sigma_1^2 = 0.185 \)
   \( \sigma_2 = 0.083 \)
   \( \rho = 0.540 \)
5. unstructured (parameter estimates shown in Table 2.)
The covariance matrices resulting from the "R" and "RCORR" options in the repeated statements are printed in Table 2. These correlation matrices, except for structure 4, "autoregressive with random effect for patient," are directly comparable with the correlation matrix in Table 1. Ignore structure 4 momentarily and compare the other correlation matrices in Table 2 with the correlation matrix in Table 1. Structures 1, "simple," and 2, "compound symmetric," clearly do not reflect the trends in Table 1. Structure 3, "autoregressive," has the general trend of correlations decreasing with length of time interval, but the values of the correlations in the autoregressive structure are too small, especially for long intervals. Thus none of the first three correlation structures appear to adequately model the correlation structure of the data. Structure 5, "unstructured," shows correlations which are very similar to the partial correlations in Table 1. The "unstructured" type is adequate, and would, in fact, be quite satisfactory for this example. Computing time can be excessive with a large number of times. The structure must be modeled by fewer parameters in order to be useful with a small number of patients.

Table 2. Covariances and Correlations from R and RCORR Options on REPEATED Statement in PROC MIXED for Five Covariance Structures

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td></td>
<td>0.267 0.0 0.0 0.0 0.0 0.0 0.0 0.0</td>
<td>0.269 0.206 0.206 0.206 0.206 0.206 0.206 0.206</td>
<td>0.266 0.228 0.195 0.167 0.143 0.123 0.105 0.090</td>
<td>0.226 0.216 0.211 0.204 0.175 0.163 0.128 0.168</td>
</tr>
<tr>
<td></td>
<td>1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0</td>
<td>1.0 0.766 0.766 0.766 0.766 0.766 0.766 0.766</td>
<td>1.0 0.856 0.733 0.629 0.538 0.461 0.394 0.338</td>
<td>0.891 0.259 0.233 0.243 0.220 0.181 0.156 0.195</td>
</tr>
<tr>
<td>(Covariance and covariances in top line, correlations in bottom line)</td>
<td>(Covariance and covariances in top line, correlations in bottom line)</td>
<td>(Covariance and covariances in top line, correlations in bottom line)</td>
<td>(Covariance and covariances in top line, correlations in bottom line)</td>
<td>(Covariances on diagonal, covariances above diagonal, correlations below diagonal)</td>
</tr>
</tbody>
</table>

Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) are indices of relative goodness of fit of estimated covariance structures. Both of these criteria are log likelihood values penalized for the number of estimated parameters. The larger the AIC or SBC values, the better the fit. AIC and SBC values for the five covariance structures are shown in Table 4. Structure 5, "unstructured," has the largest AIC, but Structure 4, "autoregressive with random effect for patient," has the largest SBC. For this example, SBC yields the best comparison of model fits because
it reflects the large number of parameters in structure 5. Based on inspection of the correlation estimates in Tables 3 and 1 and the relative values of SBC, we conclude that structure 4 "autoregressive with random effect for patient" is the best choice of covariance structure.

Table 4. Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) for Five Covariance Structures

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>simple</td>
<td>-459.5</td>
<td>-461.6</td>
</tr>
<tr>
<td>2. compound symmetric</td>
<td>-175.6</td>
<td>-179.9</td>
</tr>
<tr>
<td>3. autoregressive</td>
<td>-139.5</td>
<td>-143.8</td>
</tr>
<tr>
<td>4. autoregressive with random effect for patient</td>
<td>-126.5</td>
<td>-132.9</td>
</tr>
<tr>
<td>5. unstructured</td>
<td>-110.1</td>
<td>-187.7</td>
</tr>
</tbody>
</table>

COMPARISON OF STANDARD ERRORS OF ESTIMATES

In the previous section we observed the correlation and covariance matrices produced by five choices of structure. In this section we shall observe the effect of these choices on certain parameter estimates and their standard errors. We select a set of four comparisons among means and illustrate with ESTIMATE statements. The first three comparisons are differences between times at DRUG = A. The fourth is a comparison of drugs a and b at time 1.

```
estimate 'tim1-tim8 drg a' time 1 0 0 0 0 0 0 -1
estimate 'tim2-tim8 drg a' time 0 1 0 0 0 0 0 -1
estimate 'tim7-tim8 drg a' time 0 0 0 0 0 0 1 -1
estimate 'drga-drgb tim 1' drug 1 -1 0
estimate 'tim1-tim8 drg a' time 0 0 0 0 0 0 0 0
run;
```

Parameter estimates and standard errors from running these four ESTIMATE statements with each of the five covariance structures are in Table 5.

We first discuss the results from the first three ESTIMATE statements; that is, those giving differences between times 1 and 8, 2 and 8, and 7 and 8. It is perhaps surprising that exactly the same values of the three estimates are obtained from each of the first covariances. This will not happen in all cases if data are unbalanced, nor will it happen if polynomial trends are used to model time effects. In this example the data are balanced and time is treated as a discrete
factor. However, distinctions between the five covariance structures appear in the standard errors of the estimates. Each different structure results in different standard error estimates. Structure number 1, "simple," treats the data as if all observations were independent with the same variance. This results in equal standard error estimates for all differences between time means at the same drug and differences between drug means at the same time. These are incorrect, because this structure clearly is inappropriate. Structure number 2, "compound symmetric," acknowledges between-patient variation as being greater than within-patient variation. This results in standard errors for the first three within-patient contrasts being smaller than the standard error for the fourth between-patient contrast, which is appropriate. But compound symmetry does not accommodate different standard errors of differences between times being dependent on the length of the time interval. Structure number 3, "autoregressive," results in standard errors between times which depend on the length of the time interval. The standard error is 0.1211 for the difference between times 1 and 8, 0.1158 for the difference between times 2 and 8, etc, down to 0.0564 for the difference between times 7 and 8. If the autoregressive structure is correct, then these estimates of standard errors should be estimates of the same quantities provided by the structure number 5 ("unstructured") estimates. The structure number 5 estimates range from 0.0688 for the difference between times 1 and 8 down to 0.0556 for the difference between times 7 and 8. Thus the autoregressive estimates appear too large for long time intervals (times 1 to 8) and too small for short time intervals (times 7 to 8). Finally, we examine the standard errors provided by structure number 4, "autoregressive with random effect for patient." We see that these estimates are quite similar to the structure 5 estimates.

In conclusion, similarity of structure 4 estimates to structure 5 estimates and the fact that the Schwarz Bayesian Criterion was largest for structure 4 leads us to prefer structure 4.


REFERENCES
