Statistical Analysis of Repeated Measures Data Using SAS Procedures[1,2]

R. C. Littell*,[3], P. R. Henry†, and C. B. Ammerman†

Departments of *Statistics and †Animal Science, University of Florida, Gainesville 32611-0339

ABSTRACT: Mixed linear models were developed by animal breeders to evaluate genetic potential of bulls. Application of mixed models has recently spread to all areas of research, spurred by availability of advanced computer software. Previously, mixed model analyses were implemented by adapting fixed-effect methods to models with random effects. This imposed limitations on applicability because the covariance structure was not modeled. This is the case with PROC GLM in the SAS® System. Recent versions of the SAS System include PROC MIXED. This procedure implements random effects in the statistical model and permits modeling the covariance structure of the data. Thereby, PROC MIXED can compute efficient estimates of fixed effects and valid standard errors of the estimates. Modeling the covariance structure is especially important for analysis of repeated measures data because measurements taken close in time are potentially more highly correlated than those taken far apart in time.

Key Words: Statistical Analysis, Direct Sire Comparisons, Analysis of Variance

 J. Anim. Sci. 1998. 76:1216–1231

Introduction

The term "repeated measures" as used in this paper refers to multiple responses taken in sequence on the same experimental unit, such as an animal. Usually, the responses are taken over time, as in weekly weight measurements to establish growth curves. However, the repeated measures could be taken in spatial sequence, such as dimensions of vertebrae. The typical repeated measures experiment in animal research consists of animals randomly assigned to treatment groups, and with responses measured on each animal over a sequence of time points. Repeated measures experiments are a type of factorial experiment, with treatment and time as the two factors. They have been used commonly in animal, plant, and human research for several decades, but only in recent years have statistical and computing methodologies been available to analyze them effectively and efficiently.

The objectives of repeated measures data analysis are to examine and compare response trends over time. This can involve comparisons of treatments at specific times, or averaged over time. It also can involve comparisons of times within a treatment. These are objectives common to any factorial experiment. The feature of repeated measures experiments that requires special attention in data analysis is the correlation pattern among the responses on the same animal over time.

Methods for Analyzing Repeated Measures Data

Responses measured on the same animal are correlated because they contain a common contribution from the animal. Moreover, measures on the same animal close in time tend to be more highly correlated than measures far apart in time. Also, variances of repeated measures often change with time. These potential patterns of correlation and variation may combine to produce a complicated covariance structure of repeated measures. Special methods of statistical analysis are needed for repeated measures data because of the covariance structure. Standard regression and analysis of variance methods may produce invalid results because they require mathematical assumptions that do not hold with repeated measures data.

There are several statistical methods used for analyzing repeated measures data. Ranging from most basic to most sophisticated, these include 1) separate analyses at each time point, 2) univariate analysis of variance, 3) univariate and multivariate analyses of time contrast variables, and 4) mixed model methodology. Separate analyses at each time point do not require special methods for repeated measures but do

[1]Florida Agric. Exp. Sta. journal series no. R-05716.

[2]Presented as an invited paper at the ASAS 88th Annu. Mtg., July 1996, Rapid City, SD.

[3]To whom correspondence should be addressed: P.O. Box 11039.

Received May 12, 1997.

Accepted December 1, 1997.

not directly address the objectives of examining and comparing trends over time. The other three approaches require special methodology and software.

Development of statistical methods for repeated measures data has been an active area of research in the past two decades because of advancements in computing hardware and software. Enhancements in the SAS System (SAS, 1996) reflect the advancements in methodology and hardware. When the SAS System became available on a commercial basis in 1976, it contained the GLM procedure. This procedure enabled users to perform univariate analysis of variance but did not provide valid standard errors for most estimates. Moreover, conclusions derived from univariate analysis of variance are often invalid because the methodology does not adequately address the covariance structure of repeated measures. In 1984, the REPEATED statement was added to the GLM procedure. The results provided by the REPEATED statement were based on univariate and multivariate analyses of contrast variables computed from the repeated measures variables. This approach basically bypassed the problems of covariance structure rather than addressing them directly. The REPEATED statement enabled users to obtain statistical tests for effects involving time trends. However, the tests were inefficient and the problem of incorrect standard errors remained. In addition, missing data on even one measure of an animal caused all the data for that animal to be ignored.

In 1992, the MIXED procedure was released in the SAS System. It provided capabilities of mixed model methodology for analysis of repeated measures data. Use of mixed model methodology enabled the user to directly address the covariance structure and greatly enhanced the user's ability to analyze repeated measures data by providing valid standard errors and efficient statistical tests. In addition, some missing measures do not cause all data for an animal to be ignored.

In this paper, one example data set will be used to illustrate the four methods of analysis and their respective advantages and shortcomings. The data are from an experiment that investigated effects of several supplemental sources of dietary Mg on urinary Mg excretion in lambs. Van Ravenswaay et al. (1992) described results of a similar, but different, experiment. Treatments were a basal diet (1,275 ppm Mg) supplemented with 0, 700, 1,400, or 2,100 ppm Mg as reagent-grade anhydrous $MgSO_4$, or 1,400 ppm Mg as each of three commercial oxide sources. The seven treatments are referred to as A through G in the SAS data analysis. Five lambs were assigned to each treatment and housed in individual pens. Thus, each pen (or, equivalently, each lamb) is an experimental unit. Urinary Mg was measured for each lamb on 10 consecutive days. Data from one lamb in each of treatments A, B, and D were removed because these lambs were obvious outliers, having either excessively high or low values of urinary Mg at all days. Also, the response from one lamb in treatment A at time 7 was removed because it was an apparent outlier, with an excessively large value of urinary Mg.

The statistical analysis methods illustrated in this paper focus on treatment comparisons at specific times, treatment comparisons averaged over times, and on changes over time in specific treatments. Differences between treatments A and B are computed at individual times and averaged across times. Standard errors are computed based on each of the methods of analysis where possible. Comparison of treatments A and B illustrates the effect of the lowest experimental rate of reagent-grade anhydrous $MgSO_4$. However, choice of the difference between A and B was made only for the purpose of illustrating treatment comparisons. Comparisons involving other treatments would be implemented in similar fashion. Also, comparisons of times are illustrated with differences between the mean for each time and the mean of all subsequent times for treatment D. Standard errors are computed using each method of analysis. In most cases, SAS statements are shown that produce each of the four methods of analysis, and results are summarized. However, the intent is not to present detailed instructions on use of SAS and its interpretation. Rather, the purpose is to present an overview of the methodologies and to illustrate how they are implemented in the SAS System. Other software that implements the same methods produces similar output. Results are presented in the body of the paper without formal mathematical equations. Technical details are presented in the technical appendix.

Different SAS procedures require the data set to be organized differently. Separate analyses at each time and the GLM REPEATED statement require the data to be organized in "multivariate mode." That is, there is one row per experimental unit in the data set, and the measurements at each time are considered separate response variables. Here the measurement at time 1 is EXC1, at time 2 is EXC2, and so on. Data from the example experiment were stored in a SAS data set named SU_MULT in a multivariate mode. The univariate ANOVA and MIXED procedure require that the data be organized in "univariate mode," that is, one row per experimental unit at each time, with all urinary Mg measurements entered as values of a variable named EXC. Data from the example are stored in a SAS data set named SU_UNI in a univariate mode.

Means plotted in Figure 1 show urinary Mg content increasing over time for all treatments except A, but tending to level out after 2 to 4 d. The profile for treatment A (the basal diet) shows urinary Mg content less than for other treatments on all days after d 1. Profiles for treatments B, C, and D (sulfate source) show increases in urinary Mg corresponding to increases in dietary Mg. Profiles for treatments E, F, and G (oxide sources) show less urinary Mg than for the same level (1,400 ppm) of dietary Mg in the sulfate source.

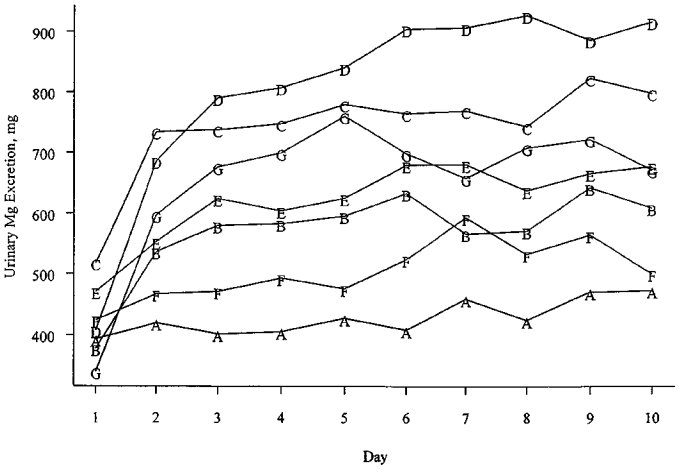

Figure 1. Urinary Mg excretions (mg) for seven diets over 10 d.

Analysis at Individual Time Points

Analysis of data at each time point examines treatment effects separately at individual observation times and makes no statistical comparisons among times. No inference is drawn about trends over time, so this method is not truly a repeated measures analysis. Use of analysis at each time point is usually at a preliminary stage of data analysis and is not a preferred method for final publication because it does not address time effects. SAS statements to obtain analyses at each time point are:

```
PROC GLM DATA = SU_MULT; CLASS TRT;
  MODEL EXC1-EXC10 = TRT;
  MEANS TRT/LSD;
  ESTIMATE 'trt A – trt B' TRT 1 –1 0 0 0 0 0;
RUN;
```

[1]

Results are summarized in Table 1, showing means for each treatment at each day, with least significant difference (**LSD**) values for comparing treatments at each day. Means and LSD values are obtained from the MEANS statement in [1]. The ESTIMATE statement in [1] computes the difference between the means for treatments A and B at each day, with standard errors, which also are presented in Table 1. Similar ESTIMATE statements could be used to compute differences between means for any treatments, or combinations of treatments. These LSD values and standard errors of estimates are valid because they only involve data at the same time.

Univariate Analysis of Variance Using the GLM Procedure

Univariate analysis of variance (ANOVA), historically, is the method most commonly applied to repeated measures data that makes comparisons between times. It treats the data as if they were from a split-plot design with the animals as whole-plot units and animals at particular times as sub-plot units. This approach also is referred to as a *split plot in time* analysis (Damon and Harvey, 1987). If measurements have equal variance at all times, and if pairs of measurements on the same animal are equally correlated, regardless of the time lag between the measurements, then the univariate ANOVA is valid from a statistical point of view, and, in fact, yields an optimal method of analysis. The condition required for validity of the univariate ANOVA tests is the so-called Huynh-Feldt (H-F) condition (Huynh and Feldt, 1970), which is mathematically less stringent than equal variances and covariances.

However, measurements close in time are often more highly correlated than measures far apart in time, which will invalidate tests for effects involving time. In the present example, tests for DAY and TRT*DAY interaction are invalid because the data fail to meet the H-F assumptions.

Table 1. Treatment means at each day, with least significant difference (LSD), and differences between means for treatments A and B, with standard errors

	Treatment									
Day	A	B	C	D	E	F	G	LSD	trt A – trt B	
1	393.4[a]	378.3	517.4	407.3	470.6	424.3[a]	340.5	202.7	15.1	(104.6)
2	420.1	536.8[a]	733.9[a]	685.6	553.5	467.0	597.7	169.5	–116.7	(87.4)
3	402.4	580.7	738.0	790.6	625.0[a]	471.4	676.7[a]	172.6	–178.3	(89.0)
4	405.7	583.5	747.7	807.9	604.9	493.4	699.7	148.5	–177.8	(76.6)
5	427.9	596.5	780.3	841.6[a]	625.6	476.4	760.7	164.9	–168.6	(85.0)
6	408.9	633.8	765.2	904.7	680.0	524.9	698.6	159.2	–225.0	(82.1)
7	459.8	567.2	769.9	907.1	681.3	594.3	658.7	207.3	–107.4	(112.3)
8	424.7	571.8	743.9	927.3	638.1	533.8	708.7	165.4	–147.1	(85.3)
9	471.2	643.3	824.9	887.4	666.8	565.7	722.7	215.4	–172.1	(111.1)
10	474.3	611.1	800.7	918.2	678.9	502.3	673.1	175.1	–136.8	(90.3)

[a]Indicates first day that is not significantly different from average of subsequent days for each treatment.

Table 2. Univariate analysis of variance

Source of variation	df	MS	Expected MS	*F*	*P*
TRT	6	729,793	$\sigma^2 + 9.95\ \sigma_d^2 + Q(trt)$	8.22	.0001
PEN(TRT)	25	88,854	$\sigma^2 + 9.96\ \sigma_d^2$		
DAY	9	188,780	$\sigma^2 + Q(day)$	21.13	.0001
TRT*DAY	54	13,067	$\sigma^2 + Q(trt*day)$	1.46	.0307
ERROR	224	8,951	σ^2		

A univariate ANOVA is obtained from PROC GLM with the following statements:

```
PROC GLM DATA = SU_UNI;
  CLASS TRT PEN DAY;
  MODEL EXC = TRT PEN(TRT) DAY TRT*DAY;
  RANDOM PEN(TRT)/TEST;
  LSMEANS TRT/STDERR E = PEN(TRT);
  LSMEANS TRT*DAY/PDIFF;
RUN;
```

[2]

Notice that the univariate version of the data set is used. The MODEL statement specifies sources of variation for the ANOVA. The RANDOM statement produces a table of expected mean squares which, in a true split-plot experiment, can be used to determine appropriate denominators of *F*-statistics for all terms in the MODEL statement. These tests are produced by the TEST option at the end of the RANDOM statement. Complete output from statements [2] is not shown. Instead, partial results are summarized in the ANOVA in Table 2. In this case, an appropriate (though not exact because of the missing datum in d 7 of treatment A) test statistic for TRT is F=MS[TRT]/MS[PEN(TRT)]. Tests for effects of DAY and TRT*DAY use *F*-statistics with MS[ERROR] for denominator mean square. These tests are not valid because the data do not satisfy the H-F condition. The first LSMEAN statement computes means for each treatment, averaged over days, with standard errors. The second LSMEANS statement computes means for combinations of treatments and days, with standard errors. Partial results from the LSMEANS statements are shown in Table 3.

In addition to the potential problems of statistical validity with univariate ANOVA analysis of repeated measures, there are potential shortcomings with capabilities of the GLM procedure. The LSMEANS statement in PROC GLM does not compute correct standard errors for the TRT*DAY means, even if correlation structure of the repeated measures is not a problem, that is, even if variances are equal and correlations are equal. For example, none of the univariate ANOVA standard errors for trt D at individual days in Table 3 are valid because they do not incorporate the variance between animals. Also, comparisons of LSMEANS between treatments at specific days (not shown) are not valid due to incorrect calculation of standard errors of differences. More detail on these problems is presented in Littell et al. (1996).

Analysis of Contrast Variables Using the GLM REPEATED Statement

Contrast variables in repeated measures data are linear combinations of the responses over time for individual animals. A familiar example from basic statistical methodology is given by the orthogonal polynomials (Snedecor and Cochran, 1980), which represent linear, quadratic, cubic, etc., trends over time. Another example is the set of differences between responses at consecutive time points, that is, changes from time 1 to time 2, time 2 to time 3, and so forth. A set of contrast variables can be used to analyze trends over time and to make comparisons between times in repeated measures data. The original repeated measures data for each animal are transformed into a new set of variables given by a set of contrast variables. Then, multivariate and univariate analyses can be applied to these new variables. This provides a method for analyzing repeated measures data that evades some of the covariance structure problems that invalidate univariate ANOVA analyses, as discussed in the previous section. The REPEATED statement in GLM provides automatic computation and analyses for several common choices of contrast variables. Data must be in a multivariate mode for use of the GLM REPEATED statement. For example, suppose that TRT is the variable indicating treatment groups and that there are measurements taken at 10 time points, entered in the data set as Y1, Y2, ... , Y10. Generic GLM statements are:

```
PROC GLM DATA = MULT;
  CLASS TRT;
  MODEL Y1–Y10 = TRT;
  REPEATED TIME <type of contrast> / <options>;
```

[3]

Note that “TIME” is not a variable in the SAS data set named MULT. Rather, it is only a name attached to the set of contrasts to be analyzed. Also, the MODEL statement in [3] is the same as the MODEL statement

Table 3. Least squares means with standard errors for treatments averaged over days and for treatment D on individual days

Mean	Univariate ANOVA	Compound symmetric	Unstructured	AR (1) plus random effects
trt A	428.9 (48.0)	428.9 (47.3)	429.7 (47.3)	428.9 (47.6)
trt B	570.3 (47.1)	570.3 (47.1)	570.3 (47.1)	570.3 (47.1)
trt C	742.2 (42.2)	742.2 (42.2)	742.2 (42.2)	742.2 (42.2)
trt D	807.8 (47.1)	807.8 (47.1)	807.8 (47.1)	807.8 (47.1)
trt E	622.5 (42.2)	622.5 (42.2)	622.5 (42.2)	622.5 (42.2)
trt F	505.3 (42.2)	505.3 (42.2)	505.3 (42.2)	505.3 (42.2)
trt G	653.7 (42.2)	653.7 (42.2)	653.7 (42.2)	653.7 (42.2)
trt D at d 1	407.3 (47.3)	407.3 (65.1)	407.3 (73.9)	407.3 (65.4)
trt D at d 2	685.6 (47.3)	685.6 (65.1)	685.6 (61.8)	685.6 (65.4)
trt D at d 3	790.7 (47.3)	790.7 (65.1)	790.7 (63.0)	790.7 (65.4)
trt D at d 4	808.0 (47.3)	808.0 (65.1)	808.0 (54.2)	808.0 (65.4)
trt D at d 5	841.6 (47.3)	841.6 (65.1)	841.6 (60.1)	841.6 (65.4)
trt D at d 6	904.7 (47.3)	904.7 (65.1)	904.7 (58.1)	904.7 (65.4)
trt D at d 7	907.1 (42.3)	907.1 (65.1)	907.1 (72.9)	907.1 (65.4)
trt D at d 8	927.3 (47.3)	927.3 (65.1)	927.3 (60.3)	927.3 (65.4)
trt D at d 9	887.4 (47.3)	887.4 (65.1)	887.4 (78.6)	887.4 (65.4)
trt D at d 10	918.2 (47.3)	918.2 (65.1)	918.2 (63.9)	918.2 (65.4)

in [1]; "TRT" is the only effect listed. Selections for "type of contrast" are listed in SAS (1989). It must be emphasized that only animals with data at all days are used in the analyses from the REPEATED statement.

Van Ravenswaay et al. (1992) used comparisons of urinary Mg at each day with the mean for subsequent days to assess when urinary Mg reached a plateau, that is, ceased to increase significantly. The HELMERT option in the REPEATED statement provides contrasts of this type. The Helmert contrasts will be used with the example data set to illustrate analyses of time contrast variables using the REPEATED statement. SAS statements to obtain these results are:

```
PROC GLM DATA=SU_MULT;
  CLASS TRT;
  MODEL EXC1–EXC10 = TRT;
  REPEATED DAY HELMERT / SUMMARY;
RUN;
```

[4]

The REPEATED statement produces results from several statistical methods to obtain tests for effects involving DAY. Partial output from the REPEATED statement in [4] is summarized in Table 4, showing *P*-values for five test statistics for DAY and TRT*DAY effects. Two multivariate tests are shown (Pillai's trace and Roy's greatest root), although the REPEATED statement actually produces four multivariate tests. Results of the two tests not shown agree approximately with those of Pillai's trace. If there were the same number of animals per treatment and no missing data on any animal, then all four multivariate tests would have equal results.

All five test statistics for DAY in Table 4 show highly significant effects of DAY ($P = .0001$). This is not a point of contention. However, there is considerable disagreement in the results for TRT*DAY. Pillai's trace finds TRT*DAY nonsignificant ($P = .3361$), whereas Roy's greatest root declares TRT*DAY highly significant ($P = .0013$). Discrepancies such as this are explainable but are common occurrences for these two multivariate tests. The REPEATED statement also produces tests that are adjustments to the *P*-value from the univariate ANOVA. They are shown in Table 4. (If all animals had complete data, the univariate ANOVA results in Table 4 would agree exactly with those in Table 2.) Based on our experiences, we prefer the G-G adjusted *P*-value instead of the multivariate tests.

Other results from the REPEATED statement in [4] are shown in Table 5. The label DAY.1 refers to a difference between the response EXC1 on d 1 and the mean of responses EXC2 on d 2 through EXC10 on d 10. That is, DAY.1 = EXC1 – (EXC2 + ... + EXC10)/9. Likewise, the label DAY.2 refers to EXC2 – (EXC3 + ... + EXC10)/8, and so forth. The REPEATED statement causes PROC GLM to compute an ANOVA for each of the contrast variables DAY.1 through DAY.9. The ANOVA for DAY.1 shows treatments are significantly different ($P = .0022$), which means that the difference between EXC1 and the mean of EXC2 through EXC10 is not the same for all treatments. Actually, this is a component of interaction between treatments and days. Treatments are not significantly different for the contrast variables DAY.2 ($P = .5228$) through DAY.9 ($P = .3043$). Interpreted at face value, this would indicate no interaction between treatments and d 2 through d 10; in other words, that profiles for all treatments are parallel from d 2 through d 10. But inspection of Figure 1 indicates profiles are not all parallel. For example, the profile for treatment D

Table 4. Significant probabilities for tests of effects involving days from REPEATED statement in PROC GLM

		Multivariate		Univariate ANOVA		
Source	df	Pillai's trace	Roy's greatest root	Unadjusted	G-G adjusted	H-F adjusted
Day	9	.0001	.0001	.0001	.0001	.0001
Trt × day	54	.3361	.0013	.0885	.1686	.1215

(sulfate source, 2,100 ppm) continues to increase following d 2, but the profile for treatment C (sulfate source, 1,400 ppm) is essentially level following d 2. Thus, it is meaningful to examine trends over days for treatments individually.

CONTRAST statements can be used in GLM to assess the significance of the variables DAY.1 through DAY.9 relative to the linear combinations defined in the CONTRAST statements. In fact, the linear combinations defined by a CONTRAST statement do not have to be "contrasts" in the sense that coefficients do not have to add to zero. Such is the case with the following statements because each CONTRAST statement actually defines a treatment mean:

```
CONTRAST 'Trt A'   INTERCEPT 1 TRT 1 0 0 0 0 0 0;
CONTRAST 'Trt B'   INTERCEPT 1 TRT 0 1 0 0 0 0 0;
CONTRAST 'Trt C'   INTERCEPT 1 TRT 0 0 1 0 0 0 0;
CONTRAST 'Trt D'   INTERCEPT 1 TRT 0 0 0 1 0 0 0;
CONTRAST 'Trt E'   INTERCEPT 1 TRT 0 0 0 0 1 0 0;
CONTRAST 'Trt F'   INTERCEPT 1 TRT 0 0 0 0 0 1 0;
CONTRAST 'Trt F'   INTERCEPT 1 TRT 0 0 0 0 0 0 1;
```
[5]

The statement CONTRAST 'Trt A' defines the mean for treatment A. Thus, the objective of the CONTRAST 'Trt A' statement is to assess whether the means of DAY.1 through DAY.9 are different from zero in treatment A. Output for CONTRAST statements [5] is summarized in Table 6. The difference between the mean for d 1 and the mean for d 2 through 10 (DAY.1) is significantly different from zero for all treatments except treatments A (basal, P = .5348) and F (oxide 1,400 ppm, P = .1270). The difference between the mean for d 2 and the mean for d 3 through d 10 (DAY.2) is not significantly different from zero for treatments A (P = .7914), B (P = .3115), C (P = .4903), and F (P = .3287), but the difference is significant for treatments D, E, and G at the .1 level. The first day that is not significantly different from the mean for subsequent days might be considered the day at which the response begins to reach a plateau. These days are indicated in Table 1.

Analyses of contrast variables as performed by the REPEATED statement in GLM are valid in the sense that equal variances and correlations are not required of measures at all times. However, results from the REPEATED statement are statistical tests, and no capability is provided by GLM for computing standard errors of comparisons. Also, only specific types of comparisons are available in the REPEATED statement.

Mixed Model Analysis Using the MIXED Procedure

As noted above, analysis of repeated measures data requires special attention to the covariance structure due to the sequential nature of the data on each animal. Procedures discussed previously either avoid the issue (analysis of contrast variables) or ignore it (univariate analysis of variance). Ignoring the covariance issues may result in incorrect conclusions from the statistical analysis. Avoiding the issues may result in inefficient analyses, which is tantamount to wasting data. The general linear mixed model allows the capability to address the issue directly by modeling the covariance structure. This capability is implemented in the MIXED procedure of the SAS System.

There are two basic steps in performing a repeated measures analysis using mixed model methodology. The first step is to model the covariance structure. The second step is to analyze time trends for treatments by estimating and comparing means.

Table 5. Summary of contrast variable ANOVA, showing mean squares and significance probabilities

Source of variation	Contrast variable								
	DAY.1	DAY.2	DAY.3	DAY.4	DAY.5	DAY.6	DAY.7	DAY.8	DAY.9
Treatment	79,196	12,574	2,651	3,283	10,791	2,179	6,394	5,609	5,105
Error	16,233	14,253	9,898	6,408	13,506	11,070	22,307	4,321	3,994
P-value for F	.0022	.5228	.9466	.7930	.5802	.9745	.9375	.2958	.3043

Table 6. Significance probabilities for tests of whether mean for each day is different from mean of subsequent days, for each treatment

Treatment	DAY.1	DAY.2	DAY.3	DAY.4	DAY.5	DAY.6	DAY.7	DAY.8	DAY.9
A	.5348	.7914	.4469	.3266	.9031	.6743	.9265	.2983	.8991
B	.0027	.3115	.6862	.6143	.8790	.5064	.5826	.1051	.3181
C	.0002	.4903	.4003	.3644	.9898	.6803	.7675	.0277	.4006
D	.0001	.0044	.0703	.0344	.2581	.9203	.9588	.4631	.3395
E	.0068	.0829	.5259	.1248	.4121	.7738	.7664	.2485	.6724
F	.1270	.3287	.2214	.2810	.2042	.6133	.3755	.9948	.0346
G	.0001	.0676	.5574	.9116	.2010	.8692	.5280	.7173	.0916

Modeling the Covariance Structure Using the RANDOM and REPEATED Statements in PROC MIXED

Measures on different animals are independent, so covariance concern is only with measures on the same animal. The covariance structure refers to variances at individual times and to correlation between measures at different times on the same animal. There are basically two aspects of the correlation. First, two measures on the same animal are correlated simply because they share common contributions from the animal. This is due to variation between animals. Second, measures on the same animal close in time are often more highly correlated than measures far apart in time. This is covariation within animals. Usually, when using PROC MIXED, the variation between animals is specified by the RANDOM statement, and covariation within animals is specified by the REPEATED statement.

Numerous structures are available as options on the REPEATED and RANDOM statements in the MIXED procedure. Three different structures will be fitted to the sheep urinary Mg data, and one will be chosen as best among the three.

First, a structure known as compound symmetry (**CS**) will be fitted. This structure specifies that measures at all times have the same variance, and that all pairs of measures on the same animal have the same correlation. The implication is that the only aspect of the covariance between repeated measures is due to the animal contribution, irrespective of proximity of time. If this structure holds, then the univariate ANOVA in Table 2 would have valid tests, although the standard errors and tests of LSMEANS from statements [2] would not necessarily be valid. Compound symmetric structure can be fitted in two ways with PROC MIXED. One way is with the RANDOM statement:

```
PROC MIXED DATA=SU_UNI;
  CLASS TRT PEN DAY;
  MODEL EXC = TRT DAY TRT*DAY;
  RANDOM PEN(TRT);
RUN;
```

[6]

This RANDOM statement specifies that there is a contribution common to all measures on the same animal, which results in equal variances at all times and equal correlations between all pairs of times. Only fixed effects are included in the PROC MIXED MODEL statement.

Statements for fitting the compound symmetric structure with the REPEATED statement are:

```
PROC MIXED DATA=SU_UNI;
  CLASS TRT PEN DAY;
  MODEL EXC = TRT DAY TRT*DAY;
  REPEATED DAY / SUB=PEN(TRT) TYPE=CS R RCORR;
RUN;
```

[7]

Here, the REPEATED statement indicates via SUB=PEN(TRT) that data are correlated on the same animal (i.e., PEN(TRT)). Partial output from statements [7] is shown in Figure 2. First, the covariance matrix is printed under the heading "R Covariance Matrix for PEN(TRT) 1 A," indicating that this is the covariance submatrix for the repeated measures from the animal in PEN 1 in TRT A. All other animals are assumed to have the same covariance matrix, although heterogeneity of variances between animals can be accommodated by the MIXED procedure. The correlation matrix is printed under the heading "Correlation Matrix for PEN(TRT) 1 A," showing correlation of .472 between all pairs of days regardless of time proximity. Under the heading "Covariance Parameter Estimates" are printed estimates of the two variance components. The between-animal (inter-animal) component is 7,993, and the within-animal(intra-animal) component is 8,950. The total variance for a measurement on a randomly chosen animal, then, is 7,993 + 8,950 = 16,943. Notice that 8,951 is the residual error mean square from the ANOVA in Table 2, essentially equal to the intra-animal variance component. The correlation between two measures on the same animal, assuming compound symmetric structure, is

$$r_{CS} = 7{,}993/(7{,}993 + 8{,}950) = .472.$$

Second, a general structure will be fitted. As an option in PROC MIXED, this is indicated as "UN" for "unstructured." This structure makes no assumptions regarding equal variances or correlations. Statements

for fitting this structure with the REPEATED statement are

```
PROC MIXED;
  CLASS TRT PEN DAY;
  MODEL EXC = TRT DAY TRT*DAY;
  REPEATED DAY / SUB=PEN(TRT) TYPE=UN R RCORR;
RUN;
```

[8]

Again, no RANDOM statement is used because inter-animal variance is absorbed into the general structure. Results from statements [8] are displayed in Figure 3, showing the covariance submatrix for PEN 1 in TRT A. All other animals have the same covariance matrix. The diagonal of this matrix shows variances ranging from 11,755 on d 4 to 24,703 on d 9. The largest variance is approximately twice as large as the smallest. This is not strong statistical evidence of unequal population variances on the different days. In particular, there is no evidence of generally increasing or decreasing trends in the variances, as might occur with other types of repeated measures, such as growth curve data.

Next, the correlation matrix is printed under the heading "R Correlation Matrix for PEN(TRT) 1 A." The correlations tend to decrease with increasing length of time interval (lag) between the days, but the trends tend to level out after a lag of 3 d rather than decrease to zero. For example, the correlation of .708 in row 2, column 3 is correlation between responses at d 2 and d 3, of lag = 1. Reading to the right in row 2, correlations decrease from .648, of lag = 2, to .233, of lag = 9. This is typical of repeated measures covariance structure, and it suggests that the pattern can be modeled mathematically. The average correlation as a function of lag, $r_{UN(MEAN)}(lag)$, is plotted in Figure 4.

There are two major potential problems with using the unstructured covariance. One, it requires estimation of a large number of variance and covariance parameters (36 in this case) and can lead to severe computational problems, especially with unbalanced data. Two, it does not exploit existence of trends in variances and covariances over time, and thus often results in erratic patterns of standard error estimates.

The trend in the correlations observed in Figures 3 and 4 can be modeled using a combination of autoregressive structure within animals and a random effect between animals. This combination structure specifies an inter-animal random effect of differences between animals, and a correlation structure within animals that decreases with increasing lag between measures. It is fitted with the MIXED procedure using both RANDOM and REPEATED statements:

```
PROC MIXED;
  CLASS TRT PEN DAY;
  MODEL EXC = TRT DAY TRT*DAY;
  RANDOM PEN(TRT);
  REPEATED DAY / SUB=PEN(TRT) TYPE=AR(1);
RUN;
```

[9]

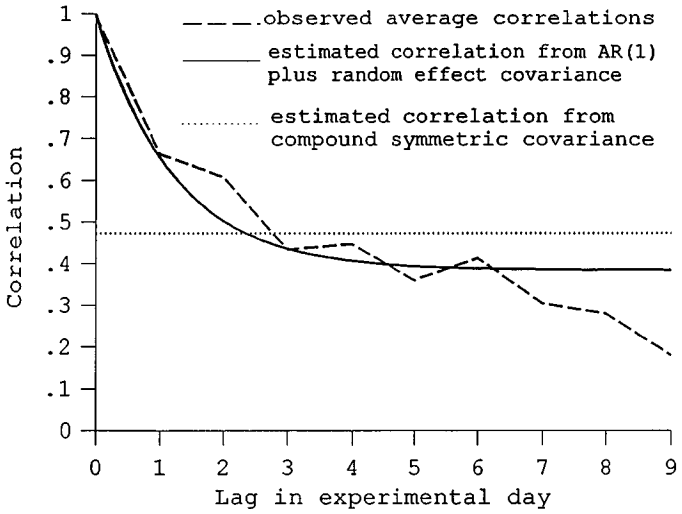

Figure 4. Observed average correlations and estimated correlation functions from compound symmetric and autoregressive plus random effect covariance structures.

Results are printed in Figure 5, showing the covariance and correlation matrices for animal 1 in treatment A. The estimates of the covariance structure parameters are 6,499 for the inter-animal variance, 10,609 for the intra-animal variance, and autoregressive correlation coefficient of .4643. The correlation function for the AR(1) plus random effect structure is

$$r_{AR(1)+RE}(lag) = (6{,}499 + 10{,}609(.4643)^{lag})/(6{,}499 + 10{,}609),$$

where lag is the length of the time interval between measures. This correlation function also is plotted in Figure 4. The agreement is good between $r_{AR(1)+RE}(lag)$ and $r_{UN}(lag)$ in Figure 4, at least until lag = 7. The values of $r_{UN}(lag)$ plotted in Figure 4 for lags of 7, 8, and 9 are based on only three, two, and one correlation values, respectively, and are not highly precise. Thus, AR(1)+RE seems to be a good choice for covariance structure based on visual assessment.

Covariance structures can be objectively compared using goodness of fit criteria that are printed by PROC MIXED, including the REML log likelihood (**REML logL**), Akaike information criterion (**AIC**), and Schwarz Bayesian criterion (**SBC**). The AIC and SBC are adjusted versions of REML logL to impose a penalty according to the number of parameters estimated. The penalty imposed by SBC is more severe than the one imposed by AIC. The SBC criterion will be used here. Its value for each of the covariance structures is shown below:

Structure	No. of parameters	SBC
CS	2	–1,573
UN	36	–1,636
AR(1)+RE	3	–1,557

In the form printed by PROC MIXED, SBC is negative for this example. The larger the value of SBC, the better the structure. The values of SBC for CS and AR(1)+RE are close, with AR(1)+RE slightly better. Thus, AR(1)+RE will be used as the covariance structure for this example.

Tests of Fixed Effects for Different Covariance Structures

The MIXED procedure prints tests for all fixed effects listed in the model statement. These tests are analogous to the univariate ANOVA *F*-tests in Table 2. Tests of fixed effects for MODEL statements in [7], [8], and [9] are shown in Table 7, with univariate ANOVA tests summarized from Table 2 for comparison. Test results for univariate ANOVA and CS covariance are very similar; they differ only because of the missing datum at d 7 in treatment A. Both would be valid if CS or H-F covariance structure assumptions held. Results of the mixed model test from UN covariance in Table 7 and the multivariate tests from Pillai's trace in Table 4 require no covariance structure assumption. The *P*-values from these tests for Trt*Day interaction are quite different in this example, partly because the multivariate test uses no data from the animal in treatment A with the missing value at d 7. In other applications in our experiences, results of these two tests usually have been more similar. The G-G adjusted *P*-value for Trt*Day in Table 4 (.1686) is in closer agreement with the AR(1)+RE covariance *P*-value in Table 7 (.0905). When all data were removed for the lamb in treatment A that had the missing value at d 7, the *P*-value for AR(1)+RE covariance changed to .1962, in even closer agreement with the G-G adjusted value.

Estimating and Comparing Means Using ESTIMATE Statements

Three types of comparisons are used to illustrate effects of covariance structure on estimates and standard errors of estimates. First is the difference between means for treatments A and B averaged over days. Second are differences between means for treatments A and B at each day. These types of comparisons are obtained by using the following statements in conjunction with PROC GLM statements [2] for the univariate ANOVA, or with PROC MIXED statements [7], [8], or [9]:

```
ESTIMATE 'trt A-B avg. over day 1–10' TRT 1 –1;
ESTIMATE 'trt A-B at day 1' TRT 1 –1 TRT*DAY
                        1 0 0 0 0 0 0 0 0 0
                  –1 0 0 0 0 0 0 0 0 0;
ESTIMATE 'trt A-B at day 2' TRT 1 –1 TRT*DAY
                        0 1 0 0 0 0 0 0 0 0
                  0 –1 0 0 0 0 0 0 0 0;
                  .
                  .
                  .
ESTIMATE 'trt A-B at day 9' TRT 1 –1 TRT*DAY
                        0 0 0 0 0 0 0 0 1 0
                  0 0 0 0 0 0 0 0 –1 0;
        ESTIMATE 'trt A-B at day 10' TRT 1 –1
                TRT*DAY 0 0 0 0 0 0 0 0 0 1
                 0 0 0 0 0 0 0 0 0 –1;      [10]
```

Third are differences between means for each day and the average of means for subsequent days for treatment D, which are obtained by using the following ESTIMATE statements with statements [2], [7], [8], or [9]:

```
ESTIMATE 'day 1 – days 2–10 in trt D' DAY –9 1
                                  1 1 1 1 1 1 1 1
 TRT*DAY          0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 0 0
                  –9 1 1 1 1 1 1 1 1 1 /
            DIVISOR=9;
ESTIMATE 'day 2 – days 3–10 in trt D' DAY 0 –8
                                  1 1 1 1 1 1 1 1
 TRT*DAY          0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 0 0
                  0 –8 1 1 1 1 1 1 1 1 /
            DIVISOR=8;
                  .
                  .
                  .
 ESTIMATE 'day 8 – days 9–10 in trt D' DAY 0 0
                              0 0 0 0 0 –2 1 1
 TRT*DAY          0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 –2 1 1 /
            DIVISOR=2;
 ESTIMATE 'DAY.9 in trt D' day 0 0 0 0 0 0 0 0 0
                                          –1 1
 TRT*DAY          0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 0 0
                  0 0 0 0 0 0 0 0 –1 1 /
            DIVISOR=1;                    [11]
```

See Littell et al. (1991) for details on writing ESTIMATE and CONTRAST statements. Results from using ESTIMATE statements [10] and [11] with GLM statements [2] and MIXED statements [7], [8], and [9] are shown in Table 8.

Different covariance structures produce different standard errors of estimates. The covariance structure

R Covariance Matrix for PEN(TRT) 1 A

Row	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
1	16943	7993	7993	7993	7993	7993	7993	7993	7993	7993
2	7993	16943	7993	7993	7993	7993	7993	7993	7993	7993
3	7993	7993	16943	7993	7993	7993	7993	7993	7993	7993
4	7993	7993	7993	16943	7993	7993	7993	7993	7993	7993
5	7993	7993	7993	7993	16943	7993	7993	7993	7993	7993
6	7993	7993	7993	7993	7993	16943	7993	7993	7993	7993
7	7993	7993	7993	7993	7993	7993	16943	7993	7993	7993
8	7993	7993	7993	7993	7993	7993	7993	16943	7993	7993
9	7993	7993	7993	7993	7993	7993	7993	7993	16943	7993
10	7993	7993	7993	7993	7993	7993	7993	7993	7993	16943

R Correlation Matrix for PEN(TRT) 1 A

Row	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
1	1.000	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472
2	0.472	1.000	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472
3	0.472	0.472	1.000	0.472	0.472	0.472	0.472	0.472	0.472	0.472
4	0.472	0.472	0.472	1.000	0.472	0.472	0.472	0.472	0.472	0.472
5	0.472	0.472	0.472	0.472	1.000	0.472	0.472	0.472	0.472	0.472
6	0.472	0.472	0.472	0.472	0.472	1.000	0.472	0.472	0.472	0.472
7	0.472	0.472	0.472	0.472	0.472	0.472	1.000	0.472	0.472	0.472
8	0.472	0.472	0.472	0.472	0.472	0.472	0.472	1.000	0.472	0.472
9	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472	1.000	0.472
10	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472	1.000

Covariance Parameter Estimates (REML)

Cov Parm	Ratio	Estimate	Std Error	Z	Pr > |Z|
DIAG CS	0.89297039	7992.5099631	2514.6760653	3.18	0.0015
Residual	1.00000000	8950.4759526	845.69897117	10.58	0.0001

Figure 2. Compound symmetric covariance and correlation matrices and covariance parameter estimates.

that provides the best fit is the appropriate one to use, although it might not result in the smallest standard errors. The AR(1)+RE structure was indicated as best among the covariance structures on the basis of the SBC criterion and the visual assessment in Figure 4. Thus, the standard errors resulting from AR(1)+RE are considered the most appropriate to report.

Table 8 shows estimates and standard errors of differences between means for treatments A and B averaged over days and at individual days, using univariate ANOVA (PROC GLM) and mixed models with CS, UN, and AR(1)+RE covariance structures (PROC MIXED). These estimates involve differences between animals. Table 8 also shows estimates and standard errors of differences between means for each day and subsequent days for treatment D, using the same various covariance structures. These estimates involve differences within animals. The selection of comparisons was made to illustrate 1) how estimates are affected by covariance structure and 2) how assumed covariance structure affects standard errors of comparisons of treatment and day means. Many of the estimates of differences in Table 8 have the same value regardless of covariance structure, although their standard errors may differ. Recall that there were four animals in each of treatments A and B, but one of the animals in treatment A had missing data at d 7.

Estimates of differences between treatments A and B on specific days in Table 8 are the same regardless of covariance structure except on d 7, because all animals used in the analysis had no missing data except at d 7. Standard errors of these estimates, however, depend on the choice of covariance structure for all days regardless of whether there are missing data. For example, except at d 7, the standard errors for estimates from univariate ANOVA are all 66.9, but from CS structure they are 92.0. This is because the univariate ANOVA standard errors (from GLM) are computed as if all observations are independent, or equivalently, as if the between-animal variance were zero. As far as standard errors from ESTIMATE

R Covariance Matrix for PEN(TRT) 1 A

Row	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
1	21862	10376	11105	8720	10212	5348	7184	4741	7591	3414
2	10376	15294	11039	8697	5755	3703	2347	4867	5277	3693
3	11105	11039	15852	11755	12612	4461	6947	6410	9084	6075
4	8720	8697	11755	11749	10606	6591	7692	6554	9781	7477
5	10212	5755	12612	10606	14466	6325	8339	5013	8427	5608
6	5348	3703	4461	6591	6325	13496	6402	8434	10163	7938
7	7184	2347	6947	7693	8339	6403	21260	7228	9303	7949
8	4741	4867	6410	6554	5013	8434	7228	14558	17189	12736
9	7591	5277	9084	9781	8427	10163	9303	17189	24703	18578
10	3414	3693	6075	7477	5608	7938	7949	12736	18578	16315

R Correlation Matrix for PEN(TRT) 1 A

Row	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
1	1.000	0.567	0.596	0.544	0.574	0.311	0.333	0.265	0.326	0.180
2	0.567	1.000	0.708	0.648	0.386	0.257	0.130	0.326	0.271	0.233
3	0.596	0.708	1.000	0.861	0.832	0.305	0.378	0.421	0.459	0.377
4	0.544	0.648	0.861	1.000	0.813	0.523	0.487	0.501	0.574	0.540
5	0.574	0.386	0.832	0.813	1.000	0.452	0.476	0.345	0.445	0.365
6	0.311	0.257	0.305	0.523	0.452	1.000	0.378	0.601	0.556	0.534
7	0.333	0.130	0.378	0.487	0.476	0.378	1.000	0.411	0.406	0.427
8	0.265	0.326	0.421	0.501	0.345	0.601	0.411	1.000	0.906	0.826
9	0.326	0.271	0.459	0.574	0.445	0.556	0.406	0.906	1.000	0.925
10	0.180	0.233	0.377	0.540	0.365	0.534	0.427	0.826	0.925	1.000

Figure 3. Unstructured covariance and correlation matrices.

statements are concerned, PEN(TRT) in statements [2] is considered a fixed effect by GLM, giving the standard error computation 66.9 = $(2*8{,}951/4)^{½}$, where 8,951 is the residual error mean square in Table 2. The CS standard errors incorporate the between-animal variance; 92.0 = $(2*(7{,}993 + 8{,}950)/4)^{-½}$, where 7,993 and 8,950 are the between- and within-animal variance component estimates, respectively, from Figure 2. The AR(1)+RE standard errors also incorporate the between-animal variance; 92.5 = $(2*(6{,}499 + 10{,}609)/4)^{½}$, where 6,499 and 10,609 are the between- and within-animal variance component estimates, respectively, from Figure 5.

The UN covariance does not model variation as a function of changes in time. This results in different standard errors for the trt A – trt B estimates on different days (Table 8), because computation of these standard errors does not exploit the phenomenon that true variation should change smoothly over days, if it changes at all. Consequently, standard errors of trt A – trt B estimates change erratically over days. By comparison, standard errors of trt A – trt B estimates are equal over days using CS (92.0), or AR(1)+RE (92.5) covariance structures, except for d 7, which had the missing value for trt A. This is reasonable because there is no evidence in the data of variance trends over experimental days.

Estimates of differences between means for individual days and means for subsequent days in treatment D are the same regardless of the assumed covariance structure, because all animals in treatment D have complete data (lower portion of Table 8). However, standard errors of the estimates are different for different assumed covariance structures. (Standard errors using univariate ANOVA and CS are virtually equal because the estimates of differences are within-animal comparisons and do not involve the between animal variance.) The standard errors follow smooth trends using univariate ANOVA, CS, or AR(1)+RE covariance structures, whereas they vary erratically for UN covariance. Standard errors using CS structure increase from 49.9 to 66.9, but standard errors for AR(1)+RE change very little over days. The discrepancies between CS and AR(1)+RE standard errors illustrate the need to select appropriate covariance structure. For example, the standard error of the difference between d 9 and d 10 assuming CS structure (66.9) is approximately 25% larger than the standard error for that difference assuming AR(1)+RE structure (53.5).

Summary and Conclusions

Mixed linear models can be implemented with either the GLM or the MIXED procedures in the SAS System. However, the GLM procedure is actually a fixed effects procedure with accessory features, such as the RANDOM statement, to make it useful for

V Covariance Matrix for PEN(TRT) 1 A

Row	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
1	17108	11424	8786	7561	6992	6728	6606	6549	6522	6510
2	11424	17108	11424	8786	7561	6992	6728	6606	6549	6522
3	8786	11424	17108	11424	8786	7561	6992	6728	6606	6549
4	7561	8786	11424	17108	11424	8786	7561	6992	6728	6606
5	6992	7561	8786	11424	17108	11424	8786	7561	6992	6728
6	6728	6992	7561	8786	11424	17108	11424	8786	7561	6992
7	6606	6728	6992	7561	8786	11424	17108	11424	8786	7561
8	6549	6606	6728	6992	7561	8786	11424	17108	11424	8786
9	6522	6549	6606	6728	6992	7561	8786	11424	17108	11424
10	6510	6522	6549	6606	6728	6992	7561	8786	11424	17108

V Correlation Matrix for PEN(TRT) 1 A

Row	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
1	1.000	0.668	0.514	0.442	0.409	0.393	0.386	0.383	0.381	0.381
2	0.668	1.000	0.668	0.514	0.442	0.409	0.393	0.386	0.383	0.381
3	0.514	0.668	1.000	0.668	0.514	0.442	0.409	0.393	0.386	0.383
4	0.442	0.514	0.668	1.000	0.668	0.514	0.442	0.409	0.393	0.386
5	0.409	0.442	0.514	0.668	1.000	0.668	0.514	0.442	0.409	0.393
6	0.393	0.409	0.442	0.514	0.668	1.000	0.668	0.514	0.442	0.409
7	0.386	0.393	0.409	0.442	0.514	0.668	1.000	0.668	0.514	0.442
8	0.383	0.386	0.393	0.409	0.442	0.514	0.668	1.000	0.668	0.514
9	0.381	0.383	0.386	0.393	0.409	0.442	0.514	0.668	1.000	0.668
10	0.381	0.381	0.383	0.386	0.393	0.409	0.442	0.514	0.668	1.000

Covariance Parameter Estimates (REML)

Cov Parm	Ratio	Estimate	Std Error	Z	Pr > |Z|
PEN(TRT)	0.61265079	6499.3990855	2620.2068465	2.48	0.0131
DIAG AR(1)	0.00004376	0.46425052	0.07797154	5.95	0.0001
Residual	1.00000000	10608.652052	1526.9861213	6.95	0.0001

Figure 5. Autoregressive order 1 plus random pen effect covariance and correlation matrices and covariance parameter estimates.

analyzing certain aspects of mixed model data. Most tests of hypothesis in an ANOVA can be computed correctly with GLM but require optional specifications, such as the TEST statement. Standard errors from LSMEANS and ESTIMATE statements in GLM usually are not computed correctly. The MIXED procedure was written from the start as a mixed model procedure and almost always makes valid computations for tests of hypothesis and standard errors of estimates.

An important application of mixed linear models is in the analysis of repeated measures data. Until recently, repeated measures data usually were analyzed by univariate ANOVA as split plot in time data, treating the experimental unit as a whole-plot and the experimental unit at a particular time as a sub-plot. This approach may be invalid because of failures for assumptions to hold concerning variances and correlations. The univariate ANOVA can be computed using the GLM procedure, but, even if variance and

Table 7. *F*-values and significance probabilities using univariate ANOVA, and for tests of fixed effects using different covariance structures in PROC MIXED

Source	df	Univariate ANOVA	Compound symmetric	Unstructured	AR(1) plus random effect
Trt	6	8.22 (.0001)	8.30 (.0001)	8.27 (.0001)	8.17 (.0001)
Day	9	21.13 (.0001)	21.09 (.0001)	18.99 (.0001)	16.07 (.0001)
Trt × day	54	1.46 (.0307)	1.46 (.0306)	1.99 (.0315)	1.31 (.0905)

Table 8. Estimates of differences between means for treatment A and B averaged over days, and on specific days, and differences between means for each day and subsequent days for treatment D, with standard errors

Estimate	Univariate ANOVA		Compound symmetric		Unstructured		AR (1) plus random effect	
trt A – trt B avg over days 1–10	–141.4	(21.4)	–141.4	(66.7)	–140.7	(66.7)	–142.2	(67.3)
trt A – trt B at d 1	15.1	(66.9)	15.1	(92.0)	15.1	(104.6)	15.1	(92.5)
trt A – trt B at d 2	–116.7	(66.9)	–116.7	(92.0)	–116.7	(87.4)	–116.7	(92.5)
trt A – trt B at d 3	–178.3	(66.9)	–178.3	(92.0)	–178.3	(89.0)	–178.3	(92.5)
trt A – trt B at d 4	–177.8	(66.9)	–177.8	(92.0)	–177.8	(76.6)	–177.8	(92.5)
trt A – trt B at d 5	–168.6	(66.9)	–168.6	(92.0)	–168.6	(85.0)	–168.6	(92.5)
trt A – trt B at d 6	–225.0	(66.9)	–225.0	(92.0)	–225.0	(82.1)	–225.0	(92.5)
trt A – trt B at d 7	–106.9	(72.8)	–106.9	(96.4)	–99.2	(107.8)	–114.5	(95.6)
trt A – trt B at d 8	–147.1	(66.9)	–147.1	(92.0)	–147.1	(85.3)	–147.1	(92.5)
trt A – trt B at d 9	–172.1	(66.9)	–172.1	(92.0)	–172.1	(111.1)	–172.1	(92.5)
trt A – trt B at d 10	–136.8	(66.9)	–136.8	(92.0)	–136.8	(90.3)	–136.8	(92.5)
d 1 – d 2–10 in trt D	444.9	(49.9)	444.9	(49.9)	444.9	(62.5)	444.9	(53.3)
d 2 – d 3–10 in trt D	187.5	(50.2)	187.5	(50.2)	187.5	(58.5)	187.5	(53.4)
d 3 – d 4–10 in trt D	94.2	(50.6)	94.2	(50.6)	94.2	(48.8)	94.2	(53.5)
d 4 – d 5–10 in trt D	89.8	(51.1)	89.8	(51.1)	89.8	(39.3)	89.8	(53.5)
d 5 – d 6–10 in trt D	67.3	(51.8)	67.3	(51.8)	67.3	(57.2)	67.3	(53.5)
d 6 – d 7–10 in trt D	5.3	(52.9)	5.3	(52.9)	5.3	(52.3)	5.3	(53.5)
d 7 – d 8–10 in trt D	3.9	(54.6)	3.9	(54.6)	3.9	(74.0)	3.9	(53.2)
d 8 – d 9–10 in trt D	–24.5	(57.9)	–24.5	(57.9)	–24.5	(32.3)	–24.5	(52.8)
d 9 – d 10 in trt D	30.8	(31.1)	30.8	(66.9)	30.8	(31.1)	30.8	(53.3)

correlation assumptions are met, the GLM procedure will not compute appropriate standard errors.

The REPEATED statement in GLM can be used to obtain tests concerning time contrasts. These can be useful, but no estimates are provided by the REPEATED statement. Also, because the GLM REPEATED statement must be run in a multivariate mode of the data set, any missing data for an animal will cause all of the data for that animal to be deleted.

Mixed model methodology, as implemented by the MIXED procedure, makes it possible to analyze repeated measures data correctly and efficiently by first modeling the variance and correlation structure of the repeated measures. Then the estimated covariance structure is used to obtain generalized least squares estimates of treatment and time differences. Choice of appropriate covariance structure may affect the computation of estimates, particularly with unbalanced data. It almost certainly will affect computation of standard errors.

Implications

Computer software is currently available that enables researchers to analyze repeated measures data using mixed model methodology. This methodology provides more valid and efficient statistical analyses of repeated measures. Implementation of this methodology requires the data analyst to model the variance and correlation structure of the data as a first step. Then, comparisons of treatments and trends over time can be analyzed.

Literature Cited

Damon, R. A., and W. R. Harvey. 1987. Experimental Design, ANOVA, and Regression. p 320. Harper and Row, New York.

Diggle, P. J. 1988. An approach to the analysis of repeated measurements. Biometrics 44:959–971.

Harville, D. A. 1977. Maximum-likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72:320–337.

Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, Ontario.

Huynh, H., and L. S. Feldt. 1970. Conditions under which mean square ratios in repeated measures designs have exact *F*-distributions. J. Am. Stat. Assoc. 65:1582–1589.

Littell, R. C., R. J. Freund, and P. C. Spector. 1991. SAS® System for Linear Models (3rd Ed.). SAS Inst. Inc., Cary, NC.

Littell, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger. 1996. SAS® System for Mixed Models. pp 31–63. SAS Inst. Inc., Cary, NC.

McLean, R. A., W. L. Sanders, and W. W. Stroup. 1991. A unified approach to mixed linear models. The Am. Stat. 45:54–63.

Milliken, G. A., and D. E. Johnson. 1992. Analysis of Messy Data, Vol. 1: Designed Experiments. Chapman and Hall, New York.

SAS. 1989. SAS/STAT User's Guide (Version 6, 4th Ed.). SAS Inst. Inc., Cary, NC.

SAS. 1996. SAS/STAT Software: Changes and Enhancements through Release 6.11. SAS Inst. Inc., Cary, NC.

Snedecor, G. W., and W. G. Cochran. 1980. Statistical Methods (7th Ed.). Iowa State University Press, Ames.

Searle, S. R. 1971. Linear Models. John Wiley & Sons, New York.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance Components. John Wiley & Sons, New York.

van Ravenswaay, R. O., R. R. Henry, C. B. Ammerman, and R. C. Littell. 1992. Relative bioavailability of magnesium sources for ruminants as measured by urinary magnesium excretion. Anim. Feed Sci. Technol. 39:13–26.

Vonesh, E. F., and V. M. Chinchilli. 1997. Linear and Nonlinear Models for the Analysis of Repeated Measures. Marcel Dekker, New York.

Technical Appendix

The MIXED procedure is based on the general linear mixed model

$$Y = X\beta + ZU + e \qquad [12]$$

where Y is a $n \times 1$ vector of observations, β is a $p \times 1$ vector of fixed, unknown parameters, X is a $n \times p$ design matrix for the fixed effects, U is a $q \times 1$ vector of unobservable random effects, Z is a $n \times q$ design matrix for the random effects, and e is a $n \times 1$ vector of residual random errors. The random vector U is assumed normally distributed with mean $E(U) = 0$ and variance $V(U) = G$, and e is assumed normally distributed with mean $E(e) = 0$ and variance $V(e) = R$. As a consequence, the observed data vector Y is normally distributed with mean $E(Y) = X\beta$ and variance $V = V(Y) = V(ZU + e) = ZGZ' + R$. This statistical model was developed for animal breeding, primarily by C. R. Henderson (Henderson, 1984). In a typical application, Y could be a vector of weaning weights of calves born to cows that were bred to a set of bulls, and the bulls were of different breeds. The fixed effects portion $X\beta$ would model the fixed effects of breed, and the ZU portion would model the random effects of bulls. The residual error e would model variation among cows.

In recent years the general linear mixed model has been used in a large variety of applications outside animal breeding. One such application is repeated measures analysis (Vonesh and Chinchilli, 1997).

The standard model for a repeated measures experiment, such as the Mg study discussed in this paper, is

$$y_{ijk} = \mu + \alpha_i + d_{ij} + \tau_k + (\alpha\tau)_{ik} + e_{ijk} \qquad [13]$$

where y_{ijk} is the response at time k on animal j in treatment group i, μ is the overall mean, α_i is a fixed effect of treatment i, d_{ij} is a random effect of animal j in treatment group i, τ_k is a fixed effect of time k, $(\alpha\tau)_{ik}$ is a fixed interaction effect of treatment i with time k, and e_{ijk} is random error at time k on animal j in treatment i. In terms of the general linear mixed model [12], the vector β contains the fixed effect parameters μ, α_i, τ_k and $(\alpha\tau)_{ik}$. The random vector U contains the between-animal random effect variables d_{ij}, and e contains the within-animal residual errors, e_{ijk}. This model is a special case of the one proposed by Diggle (1988).

The covariance structure of repeated measures involves both the d_{ij} and e_{ijk}, but the part that requires special attention is found in the e_{ijk} terms. Usually, the d_{ij} are assumed independent with variance σ_d^2. The e_{ijk} on the same animal are assumed correlated, but e_{ijk} on different animals are uncorrelated. That is,

$$\text{cov}(e_{ijk}, e_{i'j'k'}) = 0$$

if either $j \neq j'$ or $i \neq i'$. The data on animal j in treatment i are $y_{ij1}, \ldots, y_{ijt}$, where t is the number of points for each animal. The covariance between responses at times k and l on animal j in treatment i is

$$\begin{aligned}\text{cov}(y_{ijk}, y_{ijl}) &= \text{cov}(d_{ij} + e_{ijk}, d_{ij} + e_{ijl}) \\ &= V(d_{ij}) + \text{cov}(e_{ijk}, e_{ijl}) \\ &= \sigma_d^2 + \text{cov}(e_{ijk}, e_{ijl}).\end{aligned} \qquad [14]$$

The covariance structures (compound symmetric, autoregressive, etc.) may be imposed on the e_{ijk} of the same animal.

For compound symmetric covariance structure, $\text{cov}(e_{ijk}, e_{ijl}) = \sigma_b^2$ and $\text{cov}(e_{ijk}, e_{ijk}) = V(e_{ijk}) = \sigma_b^2 + \sigma^2$. Thus, from [12] covariance between two measures on the same animal is

$$\begin{aligned}\text{cov}(y_{ijk}, y_{ijl}) &= \text{cov}(d_{ij} + e_{ijk}, d_{ij} + e_{ijl}) \\ &= V(d_{ij}) + \text{cov}(e_{ijk}, e_{ijl}) \\ &= \sigma_d^2 + \sigma_b^2\end{aligned} \qquad [15]$$

and the variance of an observation is

$$\begin{aligned}V(y_{ijk}) &= V(d_{ij} + e_{ijk}) \\ &= \sigma_d^2 + \sigma_b^2 + \sigma^2\end{aligned} \qquad [16]$$

Note that σ_d^2 and σ_b^2 appear in [15] and [16] only as the sum $\sigma_d^2 + \sigma_b^2$. Thus, there is a redundancy in the variance-covariance formulation, and either σ_d^2 or σ_b^2 must be set to zero in order to be able to estimate the other. Setting $\sigma_d^2 = 0$ drops the d_{ij} terms from the model in [13]. This causes σ_b^2 to become the between-animal variance component. Setting $\sigma_b^2 = 0$ leaves σ_d^2 as the between-animal variance component and makes e_{ijk} on the same animal independent. The variance σ^2 is the within-animal variance component.

The univariate ANOVA in Table 2 (obtained from SAS statements [2]) is based on the compound symmetric covariance structure. The *F*-tests in Table 2 are valid. However, standard errors and comparisons of LSMEANS from statements [2] are not necessarily valid because the GLM procedure basically uses fixed model methodology.

Statements [6] and [7] implement compound symmetry in the covariance structure using the MIXED procedure. Statements [6] include a RANDOM statement, which defines d_{ij} in the model [13], but no REPEATED statement, which leaves the e_{ijk} on the same animal independent. Thus, statements [6] set σ_d^2

≥ 0 and $\sigma_b^2 = 0$. Statements [7], on the other hand, do not include a RANDOM statement that deletes d_{ij} from the model [13], or, equivalently, set $\sigma_d^2 = 0$. But the REPEATED statement in [7] specifies CS structure on the e_{ijk}, thus setting $\sigma_b^2 \geq 0$.

The unstructured (UN) "structure" specified no mathematical pattern on covariances among the e_{ijk} in model [13], so that the covariance between two measures on the same animal is

$$\begin{aligned} \text{cov}(y_{ijk}, y_{ijl}) &= \text{cov}(d_{ij} + e_{ijk}, d_{ij} + e_{ijl}) \\ &= V(d_{ij}) + \text{cov}(e_{ijk}, e_{ijl}) \\ &= \sigma_d^2 + \sigma_{k,l'} \end{aligned} \quad [17]$$

where $\sigma_{k,l}$ is the covariance between e_{ijk} and e_{ijl}. The variance of an observation is

$$\begin{aligned} Y(y_{ijk}) &= V(d_{ij} + e_{ijk}) \\ &= \sigma_d^2 + \sigma_{k,k}. \end{aligned} \quad [18]$$

Because no mathematical pattern is imposed on the $\sigma_{k,l}$ parameters, and because σ_d^2 always appears in the sum with a $\sigma_{k,l}$ parameter, the σ_d^2 component must be set to zero in order to be able to estimate the $\sigma_{k,l}$ parameters. Thus, SAS statements [8] contain no RANDOM statement and implement the UN structure on the e_{ijk} with the REPEATED statement.

Autoregressive (order 1) structure specifies that

$$\begin{aligned} V(e_{ijk}) &= \sigma^2 \text{ and} \\ \text{cov}(e_{ijk}, e_{ijl}) &= \sigma^2 \rho^{|k-l|} \end{aligned} \quad [19]$$

where $|k-l|$ is the lag between times k and l. Thus the variance of an observation y_{ijk} is

$$\begin{aligned} V(y_{ijk}) &= V(d_{ij} + e_{ijk}) \\ &= \sigma_d^2 + \sigma^2 \end{aligned} \quad [20]$$

and the covariance between two observations at times k and l on the same animal is

$$\begin{aligned} \text{cov}(y_{ijk}, y_{ijl}) &= \text{cov}(d_{ij} + e_{ijk}, d_{ij} + e_{ijl}) \\ &= \sigma_d^2 + \sigma^2 \rho^{|k-l|} \end{aligned} \quad [21]$$

Unlike the situation with CS and UN structure, there is no redundancy in the mathematical formulation of this covariance structure. That is, none of the parameters in [20] and [21] can be deleted without restricting the model. The between-animal variance σ_d^2 must be specified with the RANDOM statement and the within-animal covariance $\sigma^2 \rho^{|k-l|}$ must be defined with the REPEATED statement, as in statements [9].

The REPEATED statement in GLM works on the basis of linear combinations of data vectors. Conceptually, there are new sets of "variables" computed from data on each animal of the form

$$\begin{aligned} x_{ij} &= \sum_{k=1}^{t} a_k y_{ijk} \\ &= a' y_{ij} \end{aligned}$$

where $a' = (a_1,...,a_t)$ is a vector of coefficients and $y'_{ij} = (y_{ij1},...,y_{ijt})$ is the vector of repeated measures on animal j in treatment i. For example, a′ could be a vector of coefficients for a polynomial. In the sheep Mg example, the data for a given sheep are values of y_{ij1} = EXC1, y_{ij2} = EXC2, ..., y_{ij10} = EXC10 is a row of the SAS data set SU_MULT. Linear combinations implemented by the REPEATED statement in statements [4] are

DAY.1 = EXC1–(EXC2+...+EXC10)/9
DAY.2 = EXC2–(EXC3+...+EXC10)/8
.
.
.
DAY.9 = EXC9–EXC10.

Results from the REPEATED statement, such as summarized in Table 3, are simply ANOVA computations performed on DAY.1, ..., DAY.9. The ANOVA tests in Table 3 are valid regardless of the covariance structure of the original data. Likewise, results in Table 4 are from tests generated from the CONTRAST statements [5] of whether treatment means differ from zero, applied to the variables DAY.1, ..., DAY.9. (Results from CONTRAST statements [5] using GLM are equivalent to results that would be obtained with corresponding CONTRAST statements using MIXED with UN covariance, provided there were no missing data.) The REPEATED statement produces useful output in the form of tests of hypothesis and is statistically valid because covariance concerns are avoided by analyzing the linear combination variables. However, the REPEATED statement does not produce estimates of treatment means and differences between treatment means based on the linear combination variables.

The MIXED procedure works based on principles of maximum likelihood and generalized least squares applied to the model [12].

The random data vector from model [12] has expected value $E(Y)=X\beta$ and variance matrix $V = V(Y) = ZGZ' + R$. Generalized least squares methodology provides

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad [22]$$

as the best linear unbiased estimate (BLUE) of β. The variance matrix of $\hat{\beta}$ is

$$V(\hat{\beta}) = (X'V^{-1}X)^{-1}. \quad [23]$$

Thus, statistical inference about linear combinations of fixed effects is based on linear combinations of the form $L'\beta$, which are estimated by the corresponding linear combinations of $\hat{\beta}$, $L'\hat{\beta}$. The vector of estimates is BLUE of $L'\beta$, and has covariance matrix

$$V(L'\hat{\beta}) = L'(X'V^{-1}X)^{-1}L. \quad [24]$$

However, V contains unknown variance and covariance parameters.

The MIXED procedure works in two steps. The first step is to estimate variance and covariance parameters, as specified by the RANDOM and REPEATED statements. The RANDOM statement defines the ZU portion of [12] and the REPEATED statement specifies the matrix R = V(e) in model [12]. After variance and covariance estimates are obtained, they are inserted in place of the true parameter values in [22] and [24]. Then the second step is to compute test statistics and confidence intervals based on [23] and [24], using *t* and *F* distributions.

Statements [11] are examples of linear combinations of fixed effect parameters. In terms of model [13], the true mean for treatment i at time k is

$$\mu_{ik} = \mu + \alpha_i + \tau_k + (\alpha\tau)_{ik}.$$

The ESTIMATE statement in [11], labeled 'trt A–B at day 1', computes an estimate of

$$\mu_{A1} - \mu_{B1} = \alpha_A - \alpha_B + (\alpha\tau)_{A1} - (\alpha\tau)_{B1}.$$

In similar fashion, the ESTIMATE statement in [11], labeled 'DAY.1 in trt D' computes an estimate of $\mu_{D1} - (\mu_{D2} + ... + \mu_{D10})/9 = ((\alpha\tau)_{D2} + ... + (\alpha\tau)_{D10})/9$. The statements [11] can be implemented with the univariate ANOVA statements [2], but standard errors of estimates would not necessarily be valid because the sheep urinary Mg data do not have compound symmetric covariance structure.

For more information on mixed models, see Searle (1971), Harville (1977), McLean et al. (1991), Milliken and Johnson (1992), and Searle et al. (1992).