

Linear Mixed Models with Repeated Effects

Introduction and Examples Using SAS/STAT® Software

Jerry W. Davis, University of Georgia, Griffin Campus.

Introduction

Repeated measures refer to measurements taken on the same experimental unit over time or in space. Measurements taken over time often come from growth or efficacy experiments where subjects receive a treatment and their response is monitored over time. Subjects are experimental units and can be animals, plots of land or laboratory samples, for example. Repeated measures can also be taken in spatial sequence, such as root densities at increasing soil depths or measurements along an animals spine. This tutorial will focus on repeated measures taken over time, but the concepts apply to spatial examples as well.

Repeated measures can occur in any common experimental design, such as the Completely Randomized Design, Randomized Complete Block or more complicated Split and Strip-Plot designs. A basic repeated measures experiment has treatment and time as fixed, main effects. Treatment is a **between-subjects** factor because treatment levels can only change between subjects. Measurements on each subject are for the same treatment level. Time is a **within-subjects** factor because measurements on the same subject are taken at different time points within the same treatment level. Interest lies in how treatment means change (treatment effect), how treatment means change over time (time effect), and how differences in treatment means change over time (treatment by time interaction). (Littell, et al, 2006).

This is the usual interest in any two-factor experiment, so what makes a repeated measures experiment different? The difference comes from the covariance structure of the observed data. In a standard randomized block design, treatments are randomized to units (subjects) within a block. This implies that correlations between observations within a block are equal and residual errors are independent. **Within-subject** measurements tend to have correlated residual errors and the correlation often changes with time, i.e., measurements taken at adjacent time points are more correlated than those taken at time points farther apart. This results in complex covariance structures that should be modeled to give the proper standard errors for statistical tests.

Figure 1. Between-subject experimental units and within-subject measurement points.

Between-subject factor		Within-subject factor	
Rep 1		DAP	Treatment 5
Treatment 5	Treatment 6	15	3.43
Treatment 4	Treatment 1	45	1.98
Treatment 8	Treatment 2	60	1.11
Treatment 7	Treatment 3		

A standard analysis of variance, like that done by PROC GLM (SAS/STAT® Software, 2017), assumes the **within-subject** variance-covariance matrix is homogeneous¹. If this is correct, the standard analysis of variance model is appropriate. If not, the methodology should change to account for the heterogeneous variances. (Wolfinger and Chang, 1995). A common strategy is to perform separate analyses at each time point or take the average of the repeated measures and run one analysis on the average values. Another method, called a **split-plot in time**, is to test treatment effects separately from time effects using **test** statements and specifying the treatment factor's error term (see Appendix B). In 1992, SAS Institute released the MIXED procedure. It enables the analyst to model covariance structures for repeated measures data that produce correct standard errors and efficient statistical tests. (Littell, et al, 1998).

The Repeated Statement

For the MIXED procedure, options for modeling repeated effects are listed in the **repeated** statement. For an experiment with blocks, treatments and measurements over time, the **repeated** statement can be as simple as:

repeated time / type=vc subject=treatment(block);

The variable **time** represents the points in time or space in which measurements are taken. It does not need to be labeled **time** and should represent actual time or distance measurements. Using values, such as 1, 2, 3, to represent weeks 1, 3, 7, after planting, misrepresent the actual time interval. Like a graph or regression analysis, the actual intervals more accurately reflect the relationship between **Y** and **X**.

The **type** option is for specifying the covariance structure that models the correlation among the repeated measurements. In this example, I used **VC**, which stands for variance components. It is the default and assumes zero covariance between the repeated measures. PROC GLM makes the same assumption. There are many covariance structures, (more than you probably want to hear about) to model correlation, and some common ones are discussed in the next section.

The **subject** option is where one specifies the subject or experimental unit that is measured repeatedly. In a CRD experiment, the subject is often a single unit, such as a person, animal or plot; in a RCB or split-plot experiment, the subject is often an interaction term, but need not be. Subjects can be numbered sequentially, but when subjects are nested, it is more efficient to use the same numbers. Consider an example where subjects are children within two schools. It is better to number the children 1 – 4 in each school than to number them 1 – 8. The **subject** can be written as **children*school** or **children(school)**.

As always, there are many more options available for tweaking the model and computing diagnostics, such as **r** and **rcorr** for printing variance and covariance matrices. However, following this example and substituting the appropriate terms for the **type=** and **subject=** options is sufficient to analyze data from many agriculture experiments. One additional point: there must be one observation per time point for each subject to fit a repeated measures model. Otherwise, the procedure will not converge.

¹ In 1984 a REPEATED statement was added to GLM, but it couldn't model covariance structures or estimate correct standard errors. Neither my mentor nor I used it, preferring the split-plot in time method instead.

Covariance Structures

Covariance measures how much variation in one variable is explained by another variable and is used to calculate correlations. Covariance *structures* describe mathematical patterns exhibited by covariance and correlation matrices. Some covariance structures require that the measurements occur at equally spaced intervals, while others are more flexible and do not need this requirement. Model diagnostics, such as AICC, AIC, BIC and other measures are used to select the covariance structure that best fits the data. Selecting the right covariance structure is not an end unto itself. It is an intermediate step in obtaining correct tests and inference about the fixed effect means. Table 1. Lists covariance structures useful in agriculture experiments. For a complete list, see tables 78.17 and 78.18 in version 14.2 of the SAS/STAT online documentation for PROC MIXED.

Table 1. Common covariance structures for agriculture experiments. “t” equals the number of repeated measurements.

Structure	Description	Parameters	Measurement interval
VC	Variance Components	q	Unequal or equally spaced
CS	Compound Symmetry	2	Unequal or equally spaced
AR(1)	First-order Autoregressive	2	Equally spaced
SP(pow)	Spatial Power	2	Unequal or equally spaced
UN	Unstructured	$t(t+1)/2$	Unequal or equally spaced

- **VC** – As I mentioned before, variance components assumes uncorrelated errors, which is not very useful when the goal is to model the covariance to better estimate standard errors for testing means. So, if you are making the effort to model repeated effects, **do not** use **vc**. I only mention it because it is the default structure if the **type** option is not specified.
- **CS** – Compound symmetry assumes the covariance is non-zero, but is the same for all **within-subject** measurements. Correlations are the same regardless of the lag between measurements. Aside from **VC**, **CS** is the simplest structure. The assumption that all **within-subject** correlations are equal may be unrealistic for many data sets, however.
- **AR(1)** – This is a time series structure and assumes the **within-subject** correlation is a function of time that decreases toward zero with increasing lag. Measurements need to be taken at equally spaced intervals.
- **SP(pow)** – This is a spatial structure that works much like **AR(1)**, except it doesn’t require that measurements are taken at equally spaced intervals. The syntax for specifying the time or space variable is slightly different than it is for these other structures.
- **UN** – Unstructured is a very flexible structure that can theoretically fit any set of data, because it estimates a separate parameter for each element in the covariance matrix (**R matrix**). While a structure that assumes no mathematical pattern on the covariance matrix may be attractive, estimating so many parameters can lead to convergence problems. For an experiment that has five time points, the number of estimated covariance parameters is 15. The problem only worsens as **t** increases.

Symbolic mathematical notation of these covariance structures appear in Appendix C.

Repeated Measures Analysis of Variance

These program statements are for analyzing a Randomized Complete Block experiment with eight fertilizer treatments and four replications. Measurements were taken at 15, 40 and 65² days after planting (Gomez, K. and A. A. Gomez, 1984). The days after planting times are evenly spaced.

The first mixed model seminar covered random effects, LS-means, LS-mean tests and some other mixed model options, so those topics won't be covered again. The examples below only include the PROC MIXED code illustrating the use of different covariance structures. The complete program is available online and in Appendix B.

PROC MIXED Code

```
proc mixed plots=residualpanel;
```

```
  class rep treatment dap;
```

```
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
```

Treatment and dap are fixed effects and go in the **class** and **model** statements.

```
  random intercept / subject=rep;
```

Dap is the repeated measure variable. Compound symmetry is the covariance structure that models dap.

Use the more efficient syntax for the **random** statement because the model has random and repeated effects.

```
  repeated dap / type=cs subject=rep*treatment r rcorr;
```

The subject (experimental unit) is **rep*treatment**. I prefer the *var1*var2* syntax for this term. The **r** and **rcorr** options will print the estimated covariance and correlation matrices.

```
  lsmeans treatment dap;
```

```
run;
```

To estimate model parameters using a different covariance structure, change the **type** option to the desired structure. For example:

Variance components: **repeated dap / type=vc subject=rep*treatment;**

First-order autoregressive: **repeated dap / type=ar(1) subject=rep*treatment;**

² Actually, this value was unspecified. I set it at 65 to make the measurements evenly spaced.

Unstructured: **repeated dap / type=un subject=rep*treatment;**

Spatial power requires a small syntax change. The simplest change is to list the repeated (or spatial) variable after **type=sp(pow)** and omit it from between **repeated** and **/**.

Spatial power: **repeated / type=sp(pow) (dap) subject=rep*treatment;**

Output Tables

Many of the output tables have the same information as for those covered in the Random Effects seminar. Some of the following tables have information that is unique to a repeated measures analysis along with standard ANOVA and LS-means tables. The following output tables compare the differences between **type=cs** and **type=un** covariance structures.

Compound Symmetry – Model information

Model Information	
Data Set	WORK.ONE
Dependent Variable	nitrogen
Covariance Structures	Variance Components, Compound Symmetry
Subject Effects	rep, rep*treatment
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Kenward-Roger2
Degrees of Freedom Method	Kenward-Roger2

Unstructured – Model information

Model Information	
Data Set	WORK.ONE
Dependent Variable	nitrogen
Covariance Structures	Variance Components, Unstructured
Subject Effects	rep, rep*treatment
Estimation Method	REML
Residual Variance Method	None
Fixed Effects SE Method	Kenward-Roger2
Degrees of Freedom Method	Kenward-Roger2

Variance Components is for **between-subject** errors, Compound Symmetry and Unstructured are for **within-subject** errors. The subject effect of **rep** comes from the **random** statement, **rep*treatment** comes from the **repeated** statement.

Compound Symmetry – Covariance parameters

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	rep	0.01475
CS	rep*treatment	-0.00079
Residual		0.06293

Unstructured – Covariance parameters

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	rep	0.01479
UN(1,1)	rep*treatment	0.04369
UN(2,1)	rep*treatment	-0.00149
UN(2,2)	rep*treatment	0.04132
UN(3,1)	rep*treatment	-0.01104
UN(3,2)	rep*treatment	0.009231
UN(3,3)	rep*treatment	0.1005

CS estimates two parameters: rep*treatment and residual. **UN** estimates six parameters ($3*(3+1)/2$). As the number of measurement points increase, the number of parameters increase at a greater rate. Estimating many parameters can lead to convergence problems, i.e., the procedure stops and there are no tests for the fixed effects. The estimate for **rep** is associated with the **random** effect.

Compound Symmetry – Fit statistics

Fit Statistics	
-2 Res Log Likelihood	43.3
AIC (Smaller is Better)	49.3
AICC (Smaller is Better)	49.7
BIC (Smaller is Better)	47.5

Unstructured – Fit statistics

Fit Statistics	
-2 Res Log Likelihood	35.9
AIC (Smaller is Better)	49.9
AICC (Smaller is Better)	51.7
BIC (Smaller is Better)	45.6

AIC and AICC are slightly better for **CS** than for **UN**, while BIC is slightly better for **UN**. The fit statistics indicate that both models fit the data about the same.

Compound Symmetry – Covariance and correlation matrices

Estimated R Matrix for rep*treatment 1 1			
Row	Col1	Col2	Col3
1	0.06214	-0.00079	-0.00079
2	-0.00079	0.06214	-0.00079
3	-0.00079	-0.00079	0.06214

Estimated R Correlation Matrix for rep*treatment 1 1			
Row	Col1	Col2	Col3
1	1.0000	-0.01266	-0.01266
2	-0.01266	1.0000	-0.01266
3	-0.01266	-0.01266	1.0000

Unstructured – Covariance and correlation matrices

Estimated R Matrix for rep*treatment 1 1			
Row	Col1	Col2	Col3
1	0.04369	-0.00149	-0.01104
2	-0.00149	0.04132	0.009231
3	-0.01104	0.009231	0.1005

Estimated R Correlation Matrix for rep*treatment 1 1			
Row	Col1	Col2	Col3
1	1.0000	-0.03505	-0.1666
2	-0.03505	1.0000	0.1433
3	-0.1666	0.1433	1.0000

The numbers above and below the diagonal for **CS** are the same because it assumes constant correlations within subjects. **UN** is not limited by this assumption so each covariance estimate is unique. Its correlation matrix is symmetric however. This means that the estimates above and below the diagonal are like a mirror image. Note that the estimates in both **R** matrices appear in the tables labeled Covariance Parameter Estimates. I think the diagonal value in the **CS R** matrix (residual estimate) is slightly different from the table value because the matrix value does not account for **rep**.

The matrices are included just to show what two different covariance structures look like. They probably are not very useful for analyzing data from a repeated measures experiment unless one wants to take a deep dive into the analysis.

Compound Symmetry – Type 3 tests of fixed effects

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
treatment	7	21	3.47	0.0125
dap	2	48	396.07	<.0001
treatment*dap	14	48	4.28	<.0001

The two models have slightly different test statistics, but the inference derived from either model is the same.

Note that the **UN** repeated effects have fewer denominator degrees of freedom than for **CS**. This may be due to **UN** having estimated more covariance parameters. A general rule of statistical analysis is that one degree of freedom is lost for each estimated parameter. (I am not exactly sure how that applies to covariance parameters and **ddfm=kr2**.)

Unstructured – Type 3 tests of fixed effects

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
treatment	7	21.3	3.53	0.0113
dap	2	23	307.45	<.0001
treatment*dap	14	30.6	5.25	<.0001

Compound Symmetry – LS-means

Least Squares Means							
Effect	treatment	dap	Estimate	Standard Error	DF	t Value	Pr > t
treatment	1		2.0417	0.09346	10.7	21.85	<.0001
treatment	2		2.2533	0.09346	10.7	24.11	<.0001
treatment	3		2.3792	0.09346	10.7	25.46	<.0001
treatment	4		2.2725	0.09346	10.7	24.32	<.0001
treatment	5		2.1133	0.09346	10.7	22.61	<.0001
treatment	6		2.1908	0.09346	10.7	23.44	<.0001
treatment	7		2.4008	0.09346	10.7	25.69	<.0001
treatment	8		2.3817	0.09346	10.7	25.48	<.0001
dap		15	3.1706	0.07503	5.07	42.26	<.0001
dap		40	2.1819	0.07503	5.07	29.08	<.0001
dap		65	1.4100	0.07503	5.07	18.79	<.0001

Like the main effect tests, the test statistics for the two models are slightly different, but the inference is the same.

The pairwise t-tests were excluded for brevity.

Unstructured – LS-means

Least Squares Means							
Effect	treatment	dap	Estimate	Standard Error	DF	t Value	Pr > t
treatment	1		2.0417	0.09310	10.2	21.93	<.0001
treatment	2		2.2533	0.09310	10.2	24.20	<.0001
treatment	3		2.3792	0.09310	10.2	25.56	<.0001
treatment	4		2.2725	0.09310	10.2	24.41	<.0001
treatment	5		2.1133	0.09310	10.2	22.70	<.0001
treatment	6		2.1908	0.09310	10.2	23.53	<.0001
treatment	7		2.4008	0.09310	10.2	25.79	<.0001
treatment	8		2.3817	0.09310	10.2	25.58	<.0001
dap		15	3.1706	0.07115	4.01	44.56	<.0001
dap		40	2.1819	0.07063	3.54	30.89	<.0001
dap		65	1.4100	0.08269	6.79	17.05	<.0001

Selecting the Best Model

The usual strategy for selecting the best covariance structure, without *a priori* knowledge of a repeated measures process, is to fit the structures appropriate for the measurement points, (equal or unequal intervals) and compare the AICC, AIC or BIC values. (Remember, the information criterion values are relative and indicate which options provide the better fit. Not how well the model fits the data in absolute terms. The difference between values for alternate models is meaningful. The actual values are not.) Doing that for the five previously mentioned covariance structures, results in the following table.

Table 2 Fit statistics for selected covariance structures.

Structure	Type	Covariance estimate	AICC	AIC	BIC
Variance Components	VC	NA	47.5	47.3	46.1
Compound Symmetry	CS	-0.0008	49.7	49.3	47.5
First-order Autoregressive	AR(1)	0.084	49.4	49.1	47.2
Spatial (power)	SP(pow)	0.906	49.4	49.1	47.2
Unstructured	UN	NA	51.7	49.9	45.2

The small differences between information criterion values indicate that the different structures model the data equally well. AICC is probably the best criterion to use for evaluating models because it is corrected for the number of parameters in each model.

Given these results, which covariance structure would you choose and what is the next step in the analysis? I would choose **CS**. It is simple and fits as well as the more complex structures. The AICC value

for **CS** is slightly higher than for **VC** (which assumes no correlation), but it does address the issue of possibly correlated **within-subject** measurements. The researcher can assure colleagues and manuscript reviewers that the repeated measurements were modeled properly. Since the interaction between treatment and days after planting (DAP) was significant, I would go one step further and separate the treatment LS-means within each level of DAP. The SAS program for the final analysis is listed in Appendix B.

This description of the experiment and analysis might satisfy manuscript reviewers.

The experiment was a randomized complete block with four replications of eight nitrogen treatments. Measurements were taken at 15, 40 and 65 days after planting (DAP). A repeated measures analysis of variance was done using PROC MIXED (SAS/STAT® Software, 2017). There were significant interactions between treatment and DAP so tests for differences between treatment LS-means were conducted within each level of DAP.

If statistical techniques are not the focus of a presentation or journal article, details about the covariance structure may not need to be reported. Although, there is probably no harm in doing so. The author should be prepared to answer questions about it however.

Discussion

This data set is not a great example for showing how results can differ when modeling different covariance structures for a repeated measures analysis. While the results were similar, the methodology for modeling correlated measurements, assessing the model fit and selecting the best covariance structure is sound. Selecting the best covariance structure is probably the most intimidating step, but it does not need to be too difficult or confusing. As long as the procedure converges and the results look believable, it is probably an acceptable analysis. The most important thing is to address possible **within-subject** correlation. This will improve the model and let colleagues and reviewers know that the experiment was analyzed correctly.

PROC GLIMMIX Differences

In 2005 SAS Institute, release PROC GLIMMIX for fitting generalized linear mixed models. This procedure can model data from repeated measures experiments where the response variables are from continuous or discrete distributions. The developers decided to change how repeated measures are modeled by removing the **repeated** statement and using a **random** statement with the **residual** option. For example:

repeated dap / type=cs subject=rep*treatment;

is replaced by:

random dap / residual type=cs subject=rep*treatment;

Multiple **random** statements are allowed, so there is no problem modeling random and repeated effects.

PROC GLIMMIX also includes a statement for testing various attributes of a mixed model. One of which is whether a mixed model is better than a model without random and / or repeated effects. The statement syntax for making this test is:

covtest 'Is GLM OK?' glm;

Adding this statement to the GLIMMIX procedure that analyzes the example data set generates this table:

Tests of Covariance Parameters Based on the Restricted Likelihood					
Label	DF	-2 Res Log Like	ChiSq	Pr > ChiSq	Note
Is GLM OK?	2	52.8899	9.58	0.0052	MI

MI: P-value based on a mixture of chi-squares.

The probability of a greater chi-square (Pr > ChiSq) value is less than 0.05 (alpha=0.05), therefore we reject the null hypothesis that this model and a model fit by PROC GLM are the same. In other words, an analysis that models the random and repeated effects for this data set is better than an analysis that only models the fixed effects.

A SAS Software program that analyzes the example data set with PROC GLIMMIX is included in Appendix B.

References

Gbur, Edward E., Walter W. Stroup, Kevin S. McCarter, Susan Durham, Linda J. Young, Mary Christman, Mark West, Matthew Kramer. 2012. Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. Madison WI. doi: 10.2134/2012.generalized-linear-mixed-models.c2.

Gomez, Kwanchai, A., A. A. Gomez. 1984. Statistical Procedures for Agricultural Research. John Wiley and Sons, Inc.

Littell, Ramon C., George A. Millikin, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger. 2006. SAS® for Mixed Models, Second Edition. Cary, NC: SAS Institute Inc.

Littell, R. C., P. R. Henry, and C. B. Ammerman. (1998), "Statistical Analysis of Repeated Measures Data Using SAS Procedures," J. Animal Science, 76:1216-1231.

SAS/STAT Software version 14.1. 2017. SAS Institute Inc., Cary NC.

Wolfinger, Russ, Ming Chang. (1995), "Comparing the SAS® GLM and MIXED Procedures for Repeated Measures," Proceedings of the Twentieth Annual SAS Users Group Conference, Cary, NC: SAS Institute Inc.

Appendix A

Convergence

An issue that data analysts face when fitting mixed models is convergence. The procedures use maximum likelihood techniques to maximize a likelihood function. The method works by iteratively trying different parameter estimates and checking for improvement in a likelihood function. When no more improvement is found, the procedure has converged and the last set of estimates are the solution to the model. Sometimes the procedure encounters mathematical errors and fails to converge because of problems with the data, the model or both. Non-convergence is like a lack of fit test - it implies that the data does not support the model.

When a mathematical error occurs during the iterative process, the procedure aborts. It may print a message in the SASLOG or output file that a problem occurred. A common message printed in the SASLOG is:


PROC MIXED:

NOTE: An infinite likelihood is assumed in iteration 0 because of a nonpositive definite estimated R matrix for 'variable names' 'place'.

PROC GLIMMIX:

NOTE: Did not converge.

GLIMMIX may also print a message box in the results file:



Did not converge.

Here are some problems that may cause non-convergence and some possible remedies.

- If it is a repeated measures analysis, make sure there is one observation for each subject at each time point. A data coding error often causes this problem.
- Response values are very large numbers, e.g., 10,000 or 1,000,000. Divide all these values by a common factor to reduce their scale. Dividing by a constant will not affect the results, only the measurement scale.
- An over parameterized model. If there are many random effects, remove some random interaction terms to try to stabilize the variance. I never include 3-way interactions in model or random statements. Three-way interactions may be needed to identify subjects for the **type=** option however.
- The model is not specified correctly. For a standard mixed model ANOVA, always list class variables in the class statement; list fixed effects in the class and model statements; list random effects in the class and random statements; list repeated effects in the class, model and repeated statements. *These guidelines may not apply when fitting complex models.*
- If there are many subjects, combining them into similar groups (making group the subject) may be helpful. For example, instead of individual cows, maybe pens or lots can characterize a herd and there may be fewer parameters to estimate.

- Check that the values for a treatment level are not all zeros or have identical values. Maximum likelihood cannot estimate zero (to my knowledge), and treatment levels with zero variance are problematic. This usually only happens with discrete response data, like counts or binomial data, not with continuous (normal) response data.
- Change the maximum likelihood estimation method. This is more useful for GLIMMIX than for MIXED and will be discussed in an upcoming seminar.
- Sometimes the procedure will stop after reaching the maximum number of iterations. Increasing the number of iterations is rarely helpful because some underlying problem is causing the convergence failure.

Sometimes the model just will not converge - finding no solution. The analyst may need to redefine the problem or sacrifice data to salvage parts of an experiment. Be careful about manipulating the data too much and always be prepared to explain what changes were made and the rationale for making them.

SAS Institute Resources

This paper is the best reference focused on problems that can occur when fitting mixed models. It discusses steps to take when problems arise and general tips to improve efficiency and problem avoidance. It is readable and well worth the time and effort spent studying it before, during or after working with mixed models.

Kiernan, Kathleen, Jill Tao, Phil Gibbs. 2016. "Tips and Strategies for Mixed Modeling with SAS/STAT® Procedures," Proceedings of the SAS Global Forum 2016 Conference, Cary NC: SAS Institute Inc.

Paper:

<http://support.sas.com/resources/papers/proceedings16/SAS6403-2016.pdf>

For some reason the link works in MS Word 2016, but does not work correctly after converting this document to a PDF file. A copy of the paper is in the OneDrive folder as KiernanTaoGibbs2016.pdf.

Video (Recorded at SAS Global Forum 2016, about 50 minutes):

<http://sasgf16.v.sas.com/detail/video/4854897690001/sas6403---tips-and-strategies-for-mixed-modeling-with-sas:stat%C2%AE-procedures>

Appendix B

SAS programs for analyzing repeated measures data

RepeatedCS1.sas – PROC MIXED program for the final analysis using compound symmetry as the covariance structure. It includes PROC PLM and tests for LS-means.

```

/*
/ RCB - repeated measures example. Gomez and Gomez section 6.2.1 p. 258.
/ Eight treatments, three measurements 15, 40, 65 (for this example) days
/ after transplanting.
/
/ Equal intervals 15 - 40: 25 days, 40 - 65: 25 days.
/
/ Analysis with PROC MIXED type=CS
/
/ Oct. 3, 2017
/=====*/

```

```
ods html close;
ods html;
```

```
data one;
  infile datalines firstobs=2;
  input rep treatment dap nitrogen @@;
```

```

datalines;
rep treatment dap nitrogen
1 1 15 3.26 2 1 15 2.98 3 1 15 2.78 4 1 15 2.77
1 1 40 1.88 2 1 40 1.74 3 1 40 1.76 4 1 40 2.00
1 1 65 1.40 2 1 65 1.24 3 1 65 1.44 4 1 65 1.25
1 2 15 3.84 2 2 15 3.74 3 2 15 3.09 4 2 15 3.36
1 2 40 2.36 2 2 40 2.14 3 2 40 1.75 4 2 40 1.57
1 2 65 1.53 2 2 65 1.21 3 2 65 1.28 4 2 65 1.17
1 3 15 3.50 2 3 15 3.49 3 3 15 3.03 4 3 15 3.36
1 3 40 2.20 2 3 40 2.28 3 3 40 2.48 4 3 40 2.47
1 3 65 1.33 2 3 65 1.54 3 3 65 1.46 4 3 65 1.41
1 4 15 3.43 2 4 15 3.45 3 4 15 2.81 4 4 15 3.32
1 4 40 2.32 2 4 40 2.33 3 4 40 2.16 4 4 40 1.99
1 4 65 1.61 2 4 65 1.33 3 4 65 1.40 4 4 65 1.12
1 5 15 3.43 2 5 15 3.24 3 5 15 3.45 4 5 15 3.09
1 5 40 1.98 2 5 40 1.70 3 5 40 1.78 4 5 40 1.74
1 5 65 1.11 2 5 65 1.25 3 5 65 1.39 4 5 65 1.20
1 6 15 3.68 2 6 15 3.24 3 6 15 2.84 4 6 15 2.91
1 6 40 2.01 2 6 40 2.33 3 6 40 2.22 4 6 40 2.00

```

```

1 6 65 1.26 2 6 65 1.44 3 6 65 1.12 4 6 65 1.24
1 7 15 2.97 2 7 15 2.90 3 7 15 2.92 4 7 15 2.42
1 7 40 2.66 2 7 40 2.74 3 7 40 2.67 4 7 40 2.98
1 7 65 1.87 2 7 65 1.81 3 7 65 1.31 4 7 65 1.56
1 8 15 3.11 2 8 15 3.04 3 8 15 3.20 4 8 15 2.81
1 8 40 2.53 2 8 40 2.22 3 8 40 2.61 4 8 40 2.22
1 8 65 3.04 2 8 65 1.28 3 8 65 1.23 4 8 65 1.29
;

*proc print;
run;

ods graphics off;

proc mixed plots=residualpanel;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  repeated dap / type=cs subject=rep*treatment r rcorr;
  lsmeans treatment dap;
  store cs1;
run;

proc plm restore=cs1;
  lsmeans treatment dap / lines;
  slice treatment*dap / sliceby=dap lines;
run;

```

RepeatedCS_UN1.sas – PROC MIXED code that models compound symmetry and unstructured covariance structures. Prints, but doesn't test, the LS-means.

```
proc mixed plots=residualpanel;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  repeated dap / type=cs subject=rep*treatment r rcorr;
  lsmeans treatment dap;
  store cs1;
run;
```

```
proc plm restore=cs1;
  lsmeans treatment dap / lines;
run;
```

```
proc mixed plots=residualpanel;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  repeated dap / type=un subject=rep*treatment r rcorr;
  lsmeans treatment dap;
  store un1;
run;
```

```
proc plm restore=un1;
  lsmeans treatment dap / lines;
run;
```


RepeatedAll1.sas – PROC MIXED code that models each covariance structure mentioned in the handout.

```
proc mixed plots=residualpanel;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  repeated dap / type=vc subject=rep*treatment;
  lsmeans treatment dap;
run;
```

```
proc mixed plots=residualpanel;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  repeated dap / type=cs subject=rep*treatment r rcorr;
  lsmeans treatment dap;
run;
```

```
proc mixed plots=residualpanel;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  repeated dap / type=ar(1) subject=rep*treatment;
  lsmeans treatment dap;
run;
```

```
proc mixed plots=residualpanel;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  repeated / type=sp(pow) (dap) subject=rep*treatment;
  lsmeans treatment dap;
run;
```

```
proc mixed plots=residualpanel;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  repeated dap / type=un subject=rep*treatment r rcorr;
  lsmeans treatment dap;
run;
```

RepeatedGlimmix1.sas – PROC GLIMMIX code for a repeated measures analysis of variance using compound symmetry. It includes the **covtest** statement for testing if this analysis is better than an analysis using PROC GLM.

```
proc glimmix;
  class rep treatment dap;
  model nitrogen = treatment dap treatment*dap / ddfm=kr2;
  random intercept / subject=rep;
  random dap / residual type=cs subject=rep*treatment;
  covtest 'Is GLM OK?' glm;
  store cs2;
run;

proc plm restore=cs2;
  lsmeans treatment dap / lines;
  slice treatment*dap / sliceby=dap lines;
run;
```

RepeatedGLM1.sas – PROC GLM code for analyzing repeated measures data using the **split-plot in time** approach.

```
proc glm;
  class rep treatment dap;
  model nitrogen = rep treatment rep*treatment dap dap*treatment;
  test h=treatment e=rep*treatment;
  means treatment / lsd lines e=rep*treatment;
  means dap / lsd lines;
  store cs3;
run;

proc plm restore=cs3;
  lsmeans treatment dap / lines;
  slice treatment*dap / sliceby=dap lines;
run;
```

Appendix C

Covariance matrices in symbolic mathematical notation. The rows and columns represent each point in time or space measurements were taken. There were three repeated measurements in these examples.

Variance Components – Covariance elements are zero, variance estimates are on the diagonal.

$$\begin{bmatrix} \sigma_B^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix}$$

Compound Symmetry – Covariance estimates are the same, variance estimate is on the diagonal.

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

First-order Autoregressive – Correlation estimate ρ (rho) is between -1 and 1, so as the exponent increases, the estimate for ρ decreases.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Spatial Power – Similar to the above, but the distances between measurements does not need to be the same. This spatial structure is two-dimensional and also works for time values.

$$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix}$$

Unstructured – Each covariance estimate above the diagonal can be unique. This makes for a flexible structure, but estimating so many parameters can be problematic.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

Excerpts are from SAS/STAT online documentation for PROC MIXED version 14.2.