

**Random, Repeated, or both?:
Analyzing a Repeated Measures Design using Proc Mixed**

Keith Dunnigan
Venturi Partners Incorporated
(Formerly Trilogy Consulting Company)

Introduction:

The purpose of this presentation is to discuss the differences in the 3 different basic ways of analyzing a repeated measures design using proc mixed. After providing a quick overview of mixed models as fit by SAS, a basic example is presented along with its associated matrices. These matrices are then used to illustrate the differences between the 3 different ways of analyzing the design using proc mixed, namely using the random statement, the repeated statement, or both statements.

Overview:

The Mixed Model fit by PROC MIXED is a generalization of that fit by GLM. It is:

$$Y = X\beta + Zv + \varepsilon, \quad \text{where} \quad E\begin{bmatrix} v \\ \varepsilon \end{bmatrix} = 0 \quad \text{and} \quad E\begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} = 0$$

$$G = \text{Var}(v), \quad R = \text{Var}(\varepsilon), \quad \text{and} \quad V = \text{Var}(Y) = ZGZ' + R$$

This model assumes the components of ε and v are normally distributed, the independence and constant variance assumptions of the general linear model are not required. The General Linear Model is a special case of the Mixed Model with $Z = 0$ and $R = \sigma^2$. Estimation is much harder in the mixed model due to the extra parameters. A likelihood type approach is used in which G and R are first estimated using either maximum likelihood (ML) or Restricted Maximum Likelihood (REML) and the estimates put into a Generalized Least Squares (GLS) equation and the final estimators derived. The resulting equations for the coefficient estimator vectors b and v are:

$$b = (X'V^{-1}X)^{-1}X'V^{-1}y, \quad \text{and} \quad v = GZ'V^{-1}(y - Xb)$$

The sums of squares are not calculated directly by proc mixed, but the F tests are estimated directly, with the formula being:

$$F = \frac{1}{\text{rank}(L)} \left[\frac{b}{v} \right]' L' (L' \hat{C} L)^{-1} \left[\frac{b}{v} \right]$$

Where $\hat{c}$ is the estimated covariance matrix of b and v . For further details, see the references cited later.

Since Estimates were used for G and R in the derivations, b is not the Best Linear Unbiased Estimator (BLUE) of β and v is not the Best Linear Unbiased Predictor (BLUP) of v , nor do the usual T and F statistics have exact T and F distributions. Rather the estimators are described as *empirical* and are described as EBLUE and EBLUP, and the distributions as *approximate* T and F distributions.

Basic Example:

Let's assume that a simple experiment produces the following data, which has the measure repeated twice for each subject. We are interested in the effects of treatment and time on the response variable:

Subject	Treatment	Time	Response
1	1	1	5
1	1	2	7
2	1	1	6
2	1	2	8
3	2	1	10
3	2	2	12
4	2	1	11
4	2	2	13

This is a classic case of a repeated measures design and is modeled customarily by including fixed effects in the model for treatment, time, and treatment by time interaction and a random effect C for subjects:

$$Y = \mu + \alpha_i + \beta_j + \alpha_i \beta_j + C_k + \varepsilon_l$$

In SAS the syntax for this model would look like:

```
proc mixed data = dat;
  class treat time subject;
  model response = treat time treat*time;
  random subject;
run;
```

Since the mixed model equation is: $y = X\beta + Zv + \varepsilon$, the associated matrices for this example are given by:

$$\beta' = [\mu \ \alpha_1 \ \alpha_2 \ \beta_1 \ \beta_2 \ \alpha_1\beta_1 \ \alpha_1\beta_2 \ \alpha_2\beta_1 \ \alpha_2\beta_2]$$

$$v' = [C_1 \ C_2 \ C_3 \ C_4]$$

$$\varepsilon' = [e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6 \ e_7 \ e_8]$$

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

In the example given, a traditional repeated measures analyses would model the following variance components:

$$G = \text{Var}(v) = \begin{bmatrix} \sigma_\beta^2 & 0 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & 0 & \sigma_\beta^2 \end{bmatrix}$$

$$R = \text{Var}(\varepsilon) = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\varepsilon^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 \end{bmatrix}$$

which is the model fit when using proc mixed and only a random statement with no repeated statement.

Remember that given G and R for the mixed model, the variance matrix V is given by $V = ZGZ' + R$. Calculating the first half of that equation, ZGZ' gives:

$$ZGZ' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_\beta^2 & 0 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & 0 & \sigma_\beta^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_\beta^2 & 0 & 0 & 0 \\ \sigma_\beta^2 & 0 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & 0 & \sigma_\beta^2 \\ 0 & 0 & 0 & \sigma_\beta^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_\beta^2 & \sigma_\beta^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_\beta^2 & \sigma_\beta^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 \end{bmatrix}$$

Finally finishing the computation and adding in R gives $V = ZGZ' + R$:

$$V = \begin{bmatrix} \sigma_\beta^2 + \sigma_\epsilon^2 & \sigma_\beta^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_\beta^2 & \sigma_\beta^2 + \sigma_\epsilon^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\beta^2 + \sigma_\epsilon^2 & \sigma_\beta^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 + \sigma_\epsilon^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\beta^2 + \sigma_\epsilon^2 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 + \sigma_\epsilon^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\beta^2 + \sigma_\epsilon^2 & \sigma_\beta^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2 + \sigma_\epsilon^2 \end{bmatrix}$$

Thus this is the variance matrix modeled in the case when we use a random statement and no repeated statement. In general (just as in this case) the variance matrix will have a block diagonal structure so that we can describe it by using only the first block. Table 1 shows the resulting variance matrices modeled when this example is analyzed using only a repeated statement or both a repeated and a random statement.

The syntax below is used to model the variance matrix given in the third column of Table 1 (using the appropriate 'type=' option from column one). Deleting the random statement models the variance matrix given in column two.

```
proc mixed data = dat;
  class treat time subject;
  model response = treat time treat*time;
  random subject;
  repeated type=un subject=subject;
run;
```

Table 1: Repeated only or Random and Repeated for several structures

<u>Structure Chosen in Repeated Statement</u>	<u>Repeated Statement only</u> (V = R)	<u>Both Repeated and Random</u> (V = R + ZGZ')
Unstructured (type=un)	$V = R = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$	$V = \begin{bmatrix} \sigma_{11} + \sigma_{\beta}^2 & \sigma_{12} + \sigma_{\beta}^2 \\ \sigma_{12} + \sigma_{\beta}^2 & \sigma_{22} + \sigma_{\beta}^2 \end{bmatrix}$
Banded Main Diagonal (type=un(1))	$V = R = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$	$V = \begin{bmatrix} \sigma_1^2 + \sigma_{\beta}^2 & \sigma_{\beta}^2 \\ \sigma_{\beta}^2 & \sigma_2^2 + \sigma_{\beta}^2 \end{bmatrix}$
Compound Symmetry (Type=cs)	$V = R = \begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$	$V = \begin{bmatrix} \sigma^2 + \sigma_1 + \sigma_{\beta}^2 & \sigma_1 + \sigma_{\beta}^2 \\ \sigma_1 + \sigma_{\beta}^2 & \sigma^2 + \sigma_1 + \sigma_{\beta}^2 \end{bmatrix}$
Heterogeneous 1 st Order Autoregressive (Type=arh(1))	$V = R = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$	$V = \begin{bmatrix} \sigma_1^2 + \sigma_{\beta}^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 + \sigma_{\beta}^2 \end{bmatrix}$

The first column of this table shows the variance matrices fit when using the repeated statement to model the R matrix. This type of modeling asks the researcher to postulate a likely variance structure representing how the different subjects may be correlated with each other. The second column shows the modeled variance matrix that results once a random statement is added to the syntax.

We can see that the addition of the random statement does not really change the form of the variance matrix since this form can be represented by using only a repeated statement. Using only the repeated statement has the advantage that it requires one less parameter. The only benefit to adding the random statement is to identify the portion of the variance or covariance that is attributable to just the subjects. Usually estimating variance components is not a goal in a repeated measures analysis.

Summary:

When analyzing a basic repeated measures design using proc mixed there is really no need to add a random statement to the SAS Code. The random subject effect can be controlled for in the model adequately by specifying an appropriate R matrix in the repeated statement. Adding a random statement does not improve the fit of the model but reduces the power and efficiency. One benefit that can be had by adding a random statement to the repeated statement is to identify the portion of the estimated variance/covariance matrix that is attributable to just the subjects, although this is not normally a goal in repeated measures analyses.

References:

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

Latour, D., Latour, K., and Wolfinger, R.D. (1994), Getting Started with
Proc Mixed,
Software Sales and Marketing Department, SAS Institute Inc., Cary, NC.