Paper AS09

Confused with PROC MIXED? Find Out What is the Best Model.

Savithri Jajam, Covance Inc., USA
Vamshi Matta, Covance Inc., USA
Lavanya Peddibhotla, Covance Inc., USA

ABSTRACT:

The MIXED procedure is flexible and offers a wide variety of covariance structures for hierarchical and repeated measures data. Mixed Model Repeated Measures (MMRM) can be used to deal with continuous endpoints when an outcome is collected multiple times or the data is structured hierarchically, i.e. children within a classroom, children within a year group, children within a school. Selection of the correct covariance structure can be challenging and often requires trial and error to find out which model fits the best. This testing involves defining various options and providing the right type of variance-covariance structure to come up with the best model. This paper discusses different types of covariance structures and includes a macro to dynamically choose the correct approach given a pre-specified hierarchy.

INTRODUCTION:

PROC MIXED is a powerful tool which is used to analyze repeated measures and hierarchical data. PROC MIXED fits mixed linear models to data. All available observations will be used in PROC MIXED, meaning that if the data is missing at any visit or timepoint, the observation can still be included in the analysis model. No additional programming effort will be needed to impute data for missing observations at any visit or timepoint. This method considers a wide variety of covariance structures which can directly address the within-subject variability and incorporate it into a statistical model. We can use multiple models in the MIXED procedure to analyze a much wider range of linear models with more types of clinical data. This specific feature has made the MIXED procedure a desirable choice for statistical analysis. The MIXED procedure is designed for easy accessibility to a wide variety of mixed models.

The focus of this paper is to describe a few different types of covariance structure and proceeds to provide a tested macro which will select the best type if the UN type fails to converge.

Topics discussed in this paper are:

1) UN: Unstructured
2) CS: Compound Symmetry
3) AR(1): Autoregressive(1)
4) TOEP: Toeplitz

UN (UNSTRUCTURED):

The UN structure is the most “liberal” of all allowing every term to be different. It requires fitting the most parameters of any structure.

Picture 1:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}$$

CS (Compound Symmetry):

The CS structure is the well-known for split-plot designs “in the old days”. As can be seen in the table, the variances are homogeneous. There is a correlation between two separate measurements, but it is assumed that the correlation is constant regardless of how far apart the measurements are.

Picture 2:

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

AR(1) (Auto regressive):

This structure has homogeneous variances and correlations that decline exponentially with distance. In our case this means that the variability in a measurement, say white blood cell count, is constant regardless of when you measure it. It also means that two measurements that are right next to each other in time are going to be correlated (depending on the value of ρ), but that as measurements get farther and farther apart, they are less correlated.

Picture 3:

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

TOEP (Toeplitz):

The TOEP structure is similar to the AR(1) in that all measurements next to each other have the same correlation, measurements two apart have the same correlation different from the first, measurements three apart have the same correlation different from the first two, etc. However, the correlations do not necessarily have the same pattern as in the AR(1). Technically, the AR(1) is a special case of the Toeplitz.

Picture 4:

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

Simple syntax with ODS output method:

Below is the simple syntax for PROC MIXED.

```
ods output lsmeans=lsm FitStatistics=fit convergencestatus=cs Diffs=LSMDiff;

proc mixed data = indsn method = reml covtest;
  class subjid tr01pg1n avisitn
    model chg = tr01pg1n avisitn tr01pg1n*avisitn base /ddfm = kenwardroger
    solution;
  repeated avisitn/subject=subjid type=un;
  lsmeans tr01pg1n*avisitn/diff cl;
run;

ods

quit;
```

Why do we need different methods?

We must test and select one best method based on our data and requirement. Typically, this is done based upon knowledge of the visit structure, analysis variable and between-subject similarities. In this paper, the first selected method is Unstructured (UN) and if that method fails for any reason, we must pick the next best method from either the CS, TOEP or AR(1) structures. This can be done automatically through a macro created to do this job.

How do we know any method failed?

If a method is not suitable based on the available data, we will see a warning in the log, as in Picture 5. Which means we should re-visit the assumptions made when specifying the analysis. Typically within a clinical trial, the assumptions will be defined in the statistical analysis plan (SAP) and details should be provided for a backup approach where there are issues with convergence. A warning will be printed in the log.

Picture 5:

The Mixed Procedure

Model Information

Data Set	WORK.X
Dependent Variable	SLEEPMW
Covariance Structure	Unstructured
Subject Effect	PT
Estimation Method	REML
Residual Variance Method	None

Class Level Information

Class	Levels	Values
PT	12	not printed
TRTSORT	2	101 102
POOLREG	1	PORTUGAL
WEEK	9	not printed

WARNING: Unable to make hessian positive definite.

Why do we need a Macro to choose the model?

Manually checking which model works under the specified conditions based upon the available data is very time consuming. For ongoing studies, it is a good idea to have a dynamic macro in place which automatically selects the best appropriate covariance structure based upon the preferred and backup method(s) specified in the SAP. Where there are multiple runs, the chosen method can be determined each time swiftly and accurately.

How does the Macro work?

The macro discussed in this paper will check Type=UN first, and in the case it does not converge then the macro will execute the model with the 3 other covariance structures (per statistical instructions). From these structures it will select the next best model and use that to create the final output.

Macro call and parameters:

```
%mmrm (var=pchg);
```

We just need 1 macro variable to execute this macro, which is the name of the variable, percentage change from baseline (PCHG), as above. Each part of the macro is described as below.

Step 1:

As a first step, the program will check if the UN structure works by running first PROC MIXED procedure and creating a macro variable from the **convergencestatus** dataset. If the reason variable has the text **CONVERGENCE CRITERIA MET**, the UN method has converged and so the macro won't check for other models and will create the final dataset. If **convergencestatus** dataset has the text **UNABLE TO MAKE HESSIAN POSITIVE DEFINITE**, this means the UN structure did not work and the macro will proceed to the next step.

```
ods output convergencestatus=cs LSmeans=lsmnDiffs = LSMDiff FitStatistics = fit;

proc mixed data = qc1 method = reml;
  class subjid tr01pg1n avisitn;
  model &var = tr01pg1n avisitn tr01pg1n*avisitn base /ddfm = kenwardroger solution;
  repeated avisitn/subject=subjid type=un;
```

```
  lsmeans tr01pg1n*avisitn/diff cl;
run;

proc sql noprint;
  select strip(upcase(reason)) into: criteria separated by "" from work.cs;
quit;

%if %index(&criteria.,CONVERGENCE CRITERIA MET.) %then
  %do;
    ods output LSmeans=lsmn Diffs = LSMDiff FitStatistics = fit;

    proc mixed data = qc1 method = reml;
      class subjid tr01pg1n avisitn;
      model &var = tr01pg1n avisitn tr01pg1n*avisitn base /ddfm = kenwardroger
      solution;
      repeated avisitn/subject=subjid type=un;
      lsmeans tr01pg1n*avisitn/diff cl;
    run;


  %end;
```

Step 2:

In the second step, the macro again uses PROC MIXED procedure to create 3 different datasets for each model. To achieve this, a do loop is used. Once the **convergencestatus** dataset is created, the unique test will be selected from the reason variable and all 3 datasets will be stacked together. Only records with **CONVERGENCE CRITERIA MET** text will be kept, and from this, the next step will be executed.

```
%let type = CS TOEP AR;

%do ord = 1 %to 3;
  ods output convergencestatus=cs LSmeans=lsmn Diffs = LSMDiff FitStatistics = fit;

  proc mixed data = qc1 method = reml;
    class subjid tr01pg1n avisitn;
    model &var = tr01pg1n avisitn tr01pg1n*avisitn base /ddfm = kenwardroger solution;
    repeated avisitn/subject=subjid type=%scan(&type,&ord) %if &ord=3 %then(1);;
    lsmeans tr01pg1n*avisitn/diff cl;
  run;

  proc sql noprint;
    select strip(upcase(reason)) into :criteria separated by "" from work.cs;
  quit;

%end;

data data&ord;
  length type $20;
  set fit;

  if Descr='AIC (Smaller is Better)';
  criteria="&criteria";
  type="%scan(&type,&ord)";
run;

data all;
  set data1 data2 data3;

  if index(criteria, 'CONVERGENCE CRITERIA MET.');

run;
```

Step 3:

After step 2, the dataset will be sorted by the Akaike Information Criteria (AIC) to select the record with the minimum value from which to create the macro relating to the preferred structure and run PROC MIXED accordingly in order to generate correct results.

```
proc sql;
  select count(distinct criteria) into :nobs from all;
quit;

proc sort data = all;
  by criteria value type;
run;

data all;
  set all;
  by criteria value type;

  if first.criteria;
run;

proc sql noprint;
  select strip(upcase(type)) into :typemin separated by "" from work.all;
quit;

ods output LSmeans=lsmn Diffs = LSMDiff;

proc mixed data = qc1 method = reml;
  class subjid tr01pg1n avisitn;
  model &var = tr01pg1n avisitn tr01pg1n*avisitn base /ddfm = kenwardroger solution;
  repeated avisitn/subject=subjid type=&typemin;
  lsmeans tr01pg1n*avisitn/diff cl;

run;
```

CONCLUSION:

PROC MIXED is flexible and can be used with various data types as it provides wide variety of models and it includes all available data. The macro described in this paper can be adopted for as many structures as needed and proceeds to generate results based upon the pre-specified requirements. Manual can be time consuming and may lead to errors, whereas automation with the macro leads to a correctly chosen method in a short time frame.

References:

1) The MIXED Procedure (SAS/STAT Software) http://support.sas.com/techsup/faq/stat_proc/mixeproc.html.
2) Guidelines for selection the covariance structure in mixed model analysis – Chuck Kincaid, COMSYS Information Technology Services, Inc., Portage, MI.
3) Milliken and Johnson, Analysis of Messy Data Volume I: Designed Experiments, New York, Chapman & Hall, 1989.
4) Schabenberger, Mixed Model Tools in SAS/Stat®, ASA Statistical Consulting Section Roundtable Conference Call, September 2004.
5) SAS Institute Inc. 2011. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc.
6) Proc Mixed - Right Options to get Right Output Shilpa Edupganti, Sheetal Nisal, NESUG 2011

Acknowledgments:

Author would like to thank the employer Covance Inc. for the approval to present this paper at this PHUSE.

CONTACT INFORMATION:

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

Savithri Jajam, Covance Inc.
savithri.jajam@covance.com
Vamshi Matta
vamshi.matta@covance.com
Lavanya Peddibhotla, Covance Inc.
lavanya.peddibhotla@covance.com