

Confused with PROC MIXED?: Find out what is the Best Model

BY
Savithri Jajam
Vamshi Matta
Lavanya Peddibhotla

June 14th – 18th , 2021



Savithri Jajam Biography

- Savithri Jajam holds a Master's degree from Western Illinois University where she studied Masters in chemistry.
- She is currently working at Covance as Senior Principal statistical programmer.
- She has extensive CDISC experience, working with both SDTM and ADaM standards.

Introduction:

- The MIXED procedure is flexible and offers a wide variety of covariance structures. MMRM(Mixed Model Repeated Measures) used to deal with the continuous endpoints when an outcome is collected multiple times.
- Model selections is challenging and often require trial and error methods to find out which model fits the best.
- This testing involves defining various options and providing the right type of variance-covariance structure and come up with a best model and right covariates.
- This paper will discuss different type of models and a macro to choose a best model without much effort.

Introduction Cont:

- PROC MIXED is a powerful tool to analyze repeated measures data.
- All available observations will be used in PROC MIXED. Which means if data is missing at any visit or timepoint it still includes into analysis.
- This method has a wide variety of covariance structures and directly address the within-subject correlation structure and incorporate it into a statistical model.
- Especially for the analysis that includes between groups effects as well as within subject effects.
- We can use multiple models in 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 very desirable choice for statistical analysis.

Different Methods:

Focus of this paper is to describe few different types tested to select the best type if UN type fails because of insufficient data.

Topics discussed in this paper are:

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

Different Methods Cont:

Type= 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.

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

Different Methods Cont:

Type= CS (Compound Symmetry):

The CS structure is the well-known for split-plot designs “in the old days: 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.

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

Different Methods Cont:

Type= AR(1) (Auto regressive):

This structure has homogeneous variances and correlations that decline exponentially with distance. Which means that two mea that are right next to each other in time are going to be correlated.

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

Different Methods Cont:

Type= TOEP (Toeplitz):

The TOEP structure is similar to the AR(1). 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.

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

Syntax:

Simple syntax with ODS output method:

```
ods output lsmeans=lsn FitStatistics=fit convergencestatus=cs Diffs=LSMDiff;  
proc mixed data = indsn method = reml covtest;  
  class subjid tr01pg1n avisitn ldhcan tranhan;  
  model chg = tr01pg1n avisitn tr01pg1n*avisitn base ldhcan tranhan/ddfm = kenwardroger solution;  
  repeated avisitn/subject=subjid type=un;  
  lsmeans tr01pg1n*avisitn/diff cl;  
run;  
ods quit;.
```

Why we need different methods?

If any of the method fails because of insufficient data and we have to select other best method.

In this paper, UN was the first type considered and if it fails then the macro should pic the best next type from CS, TOEP and AR(1).

How do we know any method failed?

Warning will be printed in the log.

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.		

How macro works?

This macro will check Type=UN first, incase if it fails then macro will execute 3 other methods and from them it will select the best model and used to create final output by using outputs(s) created from the selected method.

Macro call and parameters:

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

We just need 1 macro variable to execute this macro. **pchg** is used in the example.

%MMRM:

```
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=un;
    lsmeans tr01pg1n*avisitn/diff cl;

run;

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

quit;
```

%MMRM:

```
%if %index(&criteria.,CONVERGENCE CRITERIA MET.) %then %do;  
  ods output LSmeans=lsmn Diffs = LSMDiff(where=(avisitn=_avisitn)) 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;
```

%MMRM:

```
%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;
```

%MMRM:

```
data data&ord;  
  length type $20;  
  set fit;  
    if Descr='AIC (Smaller is Better)';  
    criteria="&criteria";  
    type="%scan(&type,&ord)";  
    %if &ord=3 %then %do; type='AR(1)';%end;  
run;
```

%MMRM:

```
data all;  
    set data1 data2 data3;  
    if index(criteria,'CONVERGENCE CRITERIA MET.');
```

run;

```
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(uppercase(type)) into :typemin separated by "" from work.all;
```

quit;

%MMRM:

```
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:

- 1) PROC Mixed is flexible to adopt and can be used with any data as it provides wide varieties of models.
- 2) It includes all available data, so no data imputation is needed.
- 3) The macro describes in this paper can be adopted for as many methods as needed and adjust based on the requirements.
- 4) Manual checking may lead to some errors, automation with the macro leads to default next best method incase first method failed.

Questions?

Savithri Jajam

Covance, Inc.

4000 Centergreen Way # 300

Cary, NC 27513

Savithri Jajam: savithri.jajam@covance.com

Vamshi Matta: vamshi.matta@covance.com

Lavanya Peddibhotla: lavanya.peddibhotla@covance.com



Covance by Labcorp is a leading global life sciences company, which provides contract research services to the drug, medical device and diagnostics, crop protection and chemical industries. COVANCE is a registered trademark and the marketing name for Covance Inc. and its subsidiaries around the world.

www.covance.com