# How to Convert Mean Separation Output to Letter Groupings in Proc Mixed: A Tip for Statistical Analysis

Peiqiang Yu

Department of Animal and Poultry Science
College of Agriculture and Bioresources, University of Saskatchewan

Presented to
SAS User Meeting on April 28, 2010

UNIVERSITY OF SASKATCHEWAN

# Background

- Before Proc Glimmix is released in Version 9.2, we use **Proc Mixed** for statistical analysis for mixed effect models
- Such as RCBD, LSD, Split-plot, Repeated Measures…
- What is a Mixed effect model?

# Mixed Effect Model?

**Model:** To describe data from a given treatment and design structures, we usually write a model as follow:

Model: Y = {trt structure components} + {design structure components} + {error structure}.

**Mixed Effect Model:** (There are two types of mixed models)

(1) If some of the factors in the model are **fixed effects** and some are **random effects**, or

(2) If there is **more than one variance** **or** **covariance component** (in the model used to characterize the variability)

# Background

- Using **Proc GLM** (OLS) to analyse data from a Mixed Effect Model is inaccurate.
- Some Journals won't accept analysis by Proc GLM, for example, Repeated Measures… we need to use Proc Mixed

# Background

- However, using Proc Mixed for multi-treatment comparison does not produce a nice output.
- It dose not assign letters to TRT means.
- We need to assign letters for each TRT by hand.
- If  you have too many TRTs, how to do by hand??

# Background

- However, there is a program: called "**pdmix800.SAS**"
- We can use SAS software + Pdmix800.sas to assign letters for TRTs when we do multi-trt comparison.
- Use an example to show you how….

# Example

## Mixed Model Analysis of Latin Square Designs (LSD)

# An example- LSD design

- This experiment was setup to investigate the effects of **4 dietary treatment (starch levels)** (T1 to T4) on **milk production** (Y dependent variable= milk yield).

## 4 x 4 LSD

| | Cow 1 | Cow 2 | Cow 3 | Cow 4 |
|---|---|---|---|---|
| Period 1 | T1 (192) | T2 (195) | T3 (292) | T4 (249) |
| Period 2 | T2 (190) | T4 (203) | T1 (218) | T3 (210) |
| Period 3 | T3 (214) | T1 (139) | T4 (245) | T2 (163) |
| Period 4 | T4 (221) | T3 (152) | T2 (204) | T1 (134) |

# Procedures to Analyze LSD using Mixed Model

**Step 1**. Modeling the variance-covariance structures [CS, AR(1), ANTE etc ] is a first *(or preliminary)* step in the analysis of repeated measures (LSD) data using mixed model methodology.

**(The first task is to obtain a better model for the variance-covariance structure of the repeated measures on the same subject, eg. cow)**

**Step 2.** Compare AIC and BIC fit statistical analysis to find the best covariance model **(AIC =Akaike information criterion; BIC =Bayesian information criterion)**

**(The smaller AIC and BIC number, the better fit of covariance model, so check AIC and BIC number).**

**Step 3**. Using the best fit variance-covariance structure model to analyse your data and then do multi-trt comparisons.

**(We are not interested in the covariance structure *(in its own right).* Instead, we are interested in obtaining a good model for the covariance structure so that comparisons and inference about fixed effect are valid).**

**Example (LSD):** The following Program: Mixed Model Code for LSD Analysis

**Step 1**. Modeling the variance-covariance structures is a first *(or preliminary)* step in the analysis of repeated measures (eg. LSD) data using mixed model methodology.

When we use a mixed model to analyze LSD, we need to test and select the most appropriate variance-covariance assumption structures, such as

- CS (compound symmetry)
- AR(1) (first order autoregressive),

Here we test CS and AR(1) variance -covariance assumption structure:

**Example (LSD):** The following Program: Mixed Model Code for LSD Analysis

```
options nocenter pageno=1;
data one;
input  per  trt $   cow   yield;
cards;

1     T1     1     192
1     T2     2     195
1     T3     3     292
1     T4     4     249
2     T2     1     190
2     T4     2     203
2     T1     3     218
2     T3     4     210
3     T3     1     214
3     T1     2     139
3     T4     3     245
3     T2     4     163
4     T4     1     221
4     T3     2     152
4     T2     3     204
4     T1     4     134
;
```

Data step

**Example:** Mixed Model Code for LSD Analysis

**<u>The first program code</u>: test <u>Compound Symmetry (CS)</u> assumption for variance-covariance structure:**

**Program step**

```
title1 'Using Mixed model for LSD analysis: a specified repeated measure analysis';
title2 'Test 1: CS=compound symmetry assumption';

proc mixed data=one covtest cl;
class  per  trt  cow;
model yield = per  trt / ddfm=kenwardroger;
repeated per /subject = cow  type =cs    r  rcorr;
run;
```

**Test CS**

## *Variance and Covariance Assumption Structures*:

**1st variance and covariance assumption structure:**

## Compound Symmetric Assumption Structure (CS):

- **Equal variances ($\sigma^2 + \sigma_1$) on the main diagonal;**
- **Equal covariance ($\sigma_1$) on all off-diagonals**
- **Equal correlation ($\rho$).**

**Table 2. The variance and covariance matrix obtained by fitting the CS structure for the example data**

| Time | 1 (Period 1) | 2 (Period 2) | 3 (Period 3) | 4 (Period 4) |
|---|---|---|---|---|
| 1 | $\sigma^2+\sigma_1$ (=928.8) | | | |
| 2 | $\sigma_1$ (=793.62) | $\sigma^2+\sigma_1$ | | |
| 3 | $\sigma_1$ | $\sigma_1$ | $\sigma^2+\sigma_1$ | |
| 4 | $\sigma_1$ | $\sigma_1$ | $\sigma_1$ | $\sigma^2+\sigma_1$ |

**<u>The following is the 2nd program code</u>: Test First Order Autoregressive - AR(1)- assumption for variance-covariance structure to see whether it is better to fit model.**

**Program step**

```
title1 'Using Mixed model for LSD analysis: a specified repeated measure analysis';
title2 'Test 2: AR(1)= First order autoregressive assumption';

proc mixed data=one covtest cl;
class  per  trt  cow;
model yield  = per  trt  /ddfm=kenwardroger;
repeated per/subject = cow type =AR(1)     r rcorr;
run;
```

**Test AR(1)**

## 2nd Variance and covariance structure:
### "AR(1) - First Order Autoregressive Assumption Structure":

- Equal variances (time1,2,3,4: $\sigma^2$) on the main diagonal.
- Unequal covariances on the off-diagonals bands.
- Unequal correlation (with increasing time interval, correlation coefficient will reduce)

Table 3. The covariance matrix obtained by fitting the AR(1) for the example data

| Time (period) | 1 (period 1) | 2 (period 2) | 3 (period 3) | 4 (period 4) |
|---|---|---|---|---|
| 1 | $\sigma^2$ | | | |
| 2 | $\rho\sigma^2$ | $\sigma^2$ | | |
| 3 | $\rho^2\sigma^2$ | $\rho\sigma^2$ | $\sigma^2$ | |
| 4 | $\rho^3\sigma^2$ | $\rho^2\sigma^2$ | $\rho\sigma^2$ | $\sigma^2$ |

# Test other various variance-covariance assumption structures

```
repeated per/subject = cow type =AR(1)     r rcorr;
repeated per/subject = cow type =HCS     r rcorr;
repeated per/subject = cow type =HAR(1)     r rcorr;
```

# Test other various variance-covariance assumption structures

**repeated** per/**subject** = cow **type** = ????    **r rcorr**;

**Summary of characteristics of selected covariance structures**

| Structures | Number parameter | Equal spacing | Equal Variance | Equal covariance | Equal correlation |
|---|---|---|---|---|---|
| SIM | 1 | no | yes | 0 | 0 |
| CS | 2 | no | yes | yes | yes |
| AR(1) | | yes | yes | no | no |
| ANTE (1) | t+(t-1) | no | no | no | no |
| UN | (t+1)t/2 | no | no | no | no |
| SP(POW) | | no | yes | no | no |
| TOEP | t | yes | yes | no | no |
| CSH | t+1 | no | no | no | yes |
| ARH(1) | t+1 | yes | no | no | no |
| TOEPH | t+(t-1) | yes | no | no | no |

AR(1)=first order autoregressive; ANTE(1)=first order ante-dependence; TOEP=Toeplitz; SP(POW)=spatial;

# **Example (LSD):** Mixed Model Code for LSD Analysis

**Step 2**. Compare AIC and BIC fit statistical analysis to find the best covariance model after fitting different models.

**(The smaller AIC and BIC number, the better fit of covariance model).**

Comparison of the results:

| | GLM | CS | AR(1) |
|---|---|---|---|
| **Model fit info** | | | |
| **BIC** | | 91.6 | 90.1 (better) |
| **AIC** | | 90.4 | 88.9 (better) |
| **Fixed effects** | | | |
| **Period** | 0.0028 | 0.0028 | 0.0226 |
| **Treatment** | 0.0014 | 0.0014 | 0.0024 |
| **Cow*** | 0.0009 | | |
| **SEM** | 5.81 | 15.24 | 16.16 |
| **SED** | 8.22 | 8.233 | 7.37 |

- **The table show that AR(1) model is better than CS.**
- **because both AIC and BIC are smaller.**
- **The smaller AIC and BIC, the better model to fit the data.**
- **So for the above example, we need to choose AR(1) for final analysis and report the results from AR(1).**

**Example (LSD):** The following Program: Mixed Model Code for LSD Analysis

**Step 3. Using the best fit covariance structure model to analyze your data** including multi-trt comparisons.

**(we are not interested in the covariance structure *(in its own right)*. Instead, we are interested in obtaining a good model for the covariance structure so that comparisons and inference about fixed effect are valid).**

In this case we will use AR(1) structure to fit the model (Final analysis)

```
proc mixed data=one covtest cl;
class per trt cow;
model yield=per trt /ddfm=kenwardroger;
repeated per/subject = cow type =AR(1);
lsmeans trt/adjust=Tukey;
run;
```

**Multi-treatment comparison using Tukey method**

# Final results: Output

The Mixed Procedure

Type 3 Tests of Fixed Effects

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| per | 3 | 5.77 | 7.14 | 0.0226 |
| trt | 3 | 5.61 | 19.05 | 0.0024 |

**P value for Trt**

**LSmeans (adjusted means for each TRT)**

Least Squares Means

| Effect | trt | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| trt | T1 | 175.43 | 16.1584 | 3.27 | 10.86 | 0.0011 |
| trt | T2 | 189.78 | 16.1584 | 3.27 | 11.74 | 0.0009 |
| trt | T3 | 215.36 | 16.1584 | 3.27 | 13.33 | 0.0006 |
| trt | T4 | 224.68 | 16.1584 | 3.27 | 13.90 | 0.0005 |

**Output show:**

- **Differences between TRTs**
- **Also test whether TRT difference is different from 0 or not.**

Differences of Least Squares Means

| Effect | trt | _trt | Estimate | Standard Error | DF | t Value | Pr > \|t\| | Adjustment | Adj P |
|---|---|---|---|---|---|---|---|---|---|
| trt | T1 | T2 | -14.3507 | 7.3688 | 5.61 | -1.95 | 0.1028 | Tukey-Kramer | 0.3081 |
| trt | T1 | T3 | -39.9347 | 7.3688 | 5.61 | -5.42 | 0.0020 | Tukey-Kramer | 0.0079 |
| trt | T1 | T4 | -49.2487 | 7.3563 | 5.61 | -6.69 | 0.0007 | Tukey-Kramer | 0.0028 |
| trt | T2 | T3 | -25.5841 | 7.3563 | 5.61 | -3.48 | 0.0147 | Tukey-Kramer | 0.0535 |
| trt | T2 | T4 | -34.8980 | 7.3688 | 5.61 | -4.74 | 0.0038 | Tukey-Kramer | 0.0147 |
| trt | T3 | T4 | -9.3139 | 7.3688 | 5.61 | -1.26 | 0.2562 | Tukey-Kramer | 0.6159 |

## For Final Conclusions and write report.

.......

# How to get the letters for each TRT?

The effects of the treatments on milk yield are significant ($P<0.01$). The effects of period on milk yield are also significant ($P<0.05$).

Table 1. Effects of the treatments on milk yield.

| | Diets | | | | | P value | |
|---|---|---|---|---|---|---|---|
| Item | T1 | T2 | T3 | T4 | SEM | Treatment | Period |
| Milk yield | 175.4 ?? | 189.8 ?? | 215.4 ?? | 224.7 ?? | 16.16 | 0.0024 | 0.0226 |

Note: Means with the same letter in the same row are not significantly different ($P>0.05$). The multi-treatment comparisons using LSD method. SEM=standard error of mean.

# Use a program “Pdmix800.SAS”

Saxton, A.M. A macro for converting mean separation output to letter groupings in Proc Mixed. In Proc. 23rd SAS Users Group Intl., SAS Institute, Cary, NC, pp1243-1246.

If you use this program, you should put the above reference in your paper or thesis to thank this guy !

**Mixed Method**

The Mixed Procedure

Differences of Least Squares Means

| Effect | trt | _trt | Estimate | Standard Error | DF | t Value | Pr > \|t\| | Adjustment | Adj P |
|---|---|---|---|---|---|---|---|---|---|
| trt | T1 | T2 | -14.3507 | 7.3688 | 5.61 | -1.95 | 0.1028 | Tukey-Kramer | 0.3081 |
| trt | T1 | T3 | -39.9347 | 7.3688 | 5.61 | -5.42 | 0.0020 | Tukey-Kramer | 0.0079 |
| trt | T1 | T4 | -49.2487 | 7.3563 | 5.61 | -6.69 | 0.0007 | Tukey-Kramer | 0.0028 |
| trt | T2 | T3 | -25.5841 | 7.3563 | 5.61 | -3.48 | 0.0147 | Tukey-Kramer | 0.0535 |
| trt | T2 | T4 | -34.8980 | 7.3688 | 5.61 | -4.74 | 0.0038 | Tukey-Kramer | 0.0147 |
| trt | T3 | T4 | -9.3139 | 7.3688 | 5.61 | -1.26 | 0.2562 | Tukey-Kramer | 0.6159 |

# After use "Pdmix800.SAS"

Effect=trt Method=Tukey-Kramer(P<.05) Set=1

| Obs | trt | Estimate | Standard Error | Letter Group |
|---|---|---|---|---|
| 1 | T4 | 224.68 | 16.1584 | A |
| 2 | T3 | 215.36 | 16.1584 | AB |
| 3 | T2 | 189.78 | 16.1584 | BC |
| 4 | T1 | 175.43 | 16.1584 | C |

***Example (LSD):*** *The following Program: Mixed Model Code for LSD Analysis*

## **Step 4.** **Final Conclusions and write report.**

…….

The effects of the treatments on milk yield are significant (P<0.01). The effects of period on milk yield are also significant (P<0.05).

Table 1. Effects of the treatments on milk yield.

| | Diets | | | | | P value | |
|---|---|---|---|---|---|---|---|
| Item | T1 | T2 | T3 | T4 | SEM | Treatment | Period |
| Milk yield | 175.4 **c** | 189.8 **bc** | 215.4 **ab** | 224.7 **a** | 16.16 | 0.0024 | 0.0226 |

Note: Means with the same letter in the same row are not significantly different (P>0.05). The multi-treatment comparisons using LSD method. SEM=standard error of mean.

**Mixed Method**

If you want to assign letters for each treatment to show significant difference, you can use the following SAS statement + pdmix800.sas.

```
proc mixed data=one covtest cl;
class per trt cow;
model yield=per trt /ddfm=kenwardroger;
repeated per/subject = cow type =AR(1);
lsmeans trt/adjust=tukey;
```

+

```
ods output diffs=ppp
lsmeans=mmm;
ods listing exclude diffs lsmeans;
run;
%include ‘c:\pdmix800.sas';
%pdmix800(ppp,mmm,alpha=.05,sort
=yes;
```

**C:\pdmix800.sas**

```
proc mixed data=one covtest cl;
class per trt cow;
model yield=per trt /ddfm=kenwardroger;
repeated per/subject = cow type =AR(1);
lsmeans trt/adjust=tukey;
ods output diffs=ppp lsmeans=mmm;
ods listing exclude diffs lsmeans;
run;
%include 'a:pdmix800.sas';
%pdmix800(ppp,mmm,alpha=.05,sort=yes)
;
```

```
Effect=trt    Method=Tukey-Kramer(P<.05)    Set=1

                                Standard    Letter
Obs    trt     Estimate         Error       Group

 1     T4       224.68         16.1584        A
 2     T3       215.36         16.1584        AB
 3     T2       189.78         16.1584        BC
 4     T1       175.43         16.1584        C
```

ile Edit View Tools Run Solutions Window Help

Results

- Mixed: Split-plot analysis us
- Mixed: Split-plot analysis us
- Mixed: Split-plot analysis us
- GLM: Split analysis using GL
- Mixed: H0 equal fit of the 3
- Mixed: H0 equal fit of the 3
- Mixed: H0 equal fit of the 3

```
title1 'Using Mixed model for LSD analysis: a specified repeated measure analysis';
title2 'Test 1: CS=compound symmetry assumption';
proc mixed data=one covtest cl;
class per trt cow;
model yield=per trt /ddfm=kenwardroger; /* Another commonly used DDFM option is Satterthwaite
however, for repeated measures, Kenwardroger is most suitable DDFM option*/
repeated per/subject = cow type =cs r rcorr;
lsmeans trt/pdiff;
/*lsmeans trt/adjust=tukey;*/    /* BON, BONFERRONI, DUNNETT, GT2,
                                 SCHEFFE, SID, SIDAK, SIM, SIMULATE, SMM, TUKEY*/
run;

title1 'Test 2: AR(1)= First order autoregressive assumption';
proc mixed data=one covtest cl;
class per trt cow;
model yield=per trt /ddfm=kenwardroger;
repeated per/subject = cow type =AR(1) r rcorr;
lsmeans trt/pdiff;
run;

title2 'Test 2: AR(1)= First order autoregressive assump...
proc mixed data=one covtest cl;
class per trt cow;
model yield=per trt /ddfm=kenwardroger;
repeated per/subject = cow type =AR(1);
lsmeans trt/pdiff;
/*lsmeans trt/adjust=tukey;*/
ods output diffs=ppp lsmeans=mmm;
ods listing exclude diffs lsmeans;
run;
%include 'a:pdmix800.sas';
%pdmix800(ppp,mmm,alpha=.05,sort=yes);
```

**Use the best fit for final analysis**

Results | Explorer | Output - (Untitled) | Log - (Untitled) | L9-Exm1-LSD work...

\WINDOWS\system32    Ln 46, Col 1

# Final results: Output

Type 3 Tests of Fixed Effects

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| per | 3 | 5.77 | 7.14 | 0.0226 |
| trt | 3 | 5.61 | 19.05 | 0.0024 |

**Effect of Trt and Period on Milk yield are significant (P<0.05) !**

Least Squares Means

| Effect | trt | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| trt | T1 | 175.43 | 16.1584 | 3.27 | 10.86 | 0.0011 |
| trt | T2 | 189.78 | 16.1584 | 3.27 | 11.74 | 0.0009 |
| trt | T3 | 215.36 | 16.1584 | 3.27 | 13.33 | 0.0006 |
| trt | T4 | 224.68 | 16.1584 | 3.27 | 13.90 | 0.0005 |

Effect=trt Method=Tukey-Kramer(P<.05) Set=1

| Obs | trt | Estimate | Standard Error | Letter Group |
|---|---|---|---|---|
| 1 | T4 | 224.68 | 16.1584 | A |
| 2 | T3 | 215.36 | 16.1584 | AB |
| 3 | T2 | 189.78 | 16.1584 | BC |
| 4 | T1 | 175.43 | 16.1584 | C |

**Multi-Trt Mean comparisons**

# Thanks !

**Peiqiang Yu**

**Email: peiqiang.yu@usask.ca**
**Tel: 966 4132**
**Room: 6D10**

UNIVERSITY OF
SASKATCHEWAN