**บทความวิชาการ**

# Balanced Two-Way Mixed Model Using SAS and Minitab

วรางคณา กีรติวิบูลย์[1*]
Warangkhana Keerativibool[1*]

## Introduction

In an experimental design, the factors affecting the experimental units can be divided into two categories: quantitative and qualitative. The different levels of factors whether there are qualitative or quantitative may appear as fixed or random. However, the level of the factors is fixed or random depending on the selection of levels for factors according to the purpose of the experiment. Therefore, it can be said that the experimental model can be classified into three types, which are fixed, random and mixed effects models. A fixed effects model is a statistical model that represents the observed quantities in terms of explanatory variables that are treated as if the quantities were non-random. This is in contrast to random effects model and mixed model in which either all or some of the explanatory variables are treated as if they arise from the random causes. A random effects model, also called a variance components model, is a kind of hierarchical linear model whose levels of factors are random from a large number of possible levels. A statistical model that containing both fixed and random effects is called a mixed model. This model is useful in a variety of fields such as biology, chemistry, physics, social sciences and branches of agriculture. It is particularly useful in settings where repeated measurements are made on the same statistical units, or where measurements are made on clusters of related statistical units. It is important to note that two versions of mixed model exist. They are called the restricted and unrestricted models. Assumptions of both versions are slightly different in terms of the random components [1]. In the case of unrestricted model, a random factor is allowed to have fixed cross factors amongst its components of variation estimated in the population whereas the restricted model is not allowed [2]. In this review article, we give a basic introduction of a balanced two-way mixed model for both restricted and unrestricted cases. Our main focus is to demonstrate how to use different procedures in SAS and Minitab to analyze such data.

[1] Department of Mathematics and Statistics, Faculty of Science, Thaksin University, Phatthalung, and Centre of Excellence in Mathematics, CHE, Bangkok.

* Corresponding author: E-mail: warang27@gmail.com

## Balanced Two-Way Mixed Model

The linear statistical model for balanced two-way mixed model, where one of the factor A is fixed with only "a" levels, the other B is random with "b" levels from the entire population and the replications within each cell are equal to "n", can be given by [1],

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \quad (1)$$

where i=1, 2, ..., a , j=1, 2, ..., b , and k=1, 2, ..., n . The usual assumptions of a restricted mixed model are as follows [3-4],

(a) $\mu$ is an overall mean,

(b) the $\tau_i$ 's are fixed effects of factor A, subject to the restriction that they are sum to zero, i.e., $\sum_{i=1}^{a}\tau_i = 0$,

(c) the $\beta_j$ 's are random effects of factor B, which are independent and identically distributed (i.i.d.) as normal $(0, \sigma^2_\beta)$,

(d) the $(\tau\beta)_{ij}$ 's are interaction effects of fixed factor A and random factor B, which are assumed to be random effects and distributed as normal $\left(0, \frac{a-1}{a}\sigma^2_{\tau\beta}\right)$ subject to the restriction that for each j their sum over fixed effects i is equals to zero,i.e., $\sum_{i=1}^{a}(\tau\beta)_{ij} = 0$. This restriction implies that certain interaction elements for different levelsof fixed factor A at the same level of random factor B are not independent and also produce a negatively correlated interaction, i.e., $\mathrm{Cov}\left[(\tau\beta)_{ij}, (\tau\beta)_{pj}\right] = -\frac{\sigma^2_{\tau\beta}}{a}$ for $i \neq p$,

(e) the $\varepsilon_{ijk}$ 's are random errors which i.i.d. normal $(0, \sigma^2)$,

(f) the $\beta_j$ 's and $\varepsilon_{ijk}$ 's are all mutually independent, and independent of the $(\tau\beta)_{ij}$ 's,

The assumptions of an unrestricted mixed model are the same as above except that:

(g) the summation restrictions in (b) and (d) are dropped,

(h) the variance of the $(\tau\beta)_{ij}$ 's in (d) is taken to be $\sigma^2_{\tau\beta}$ instead of $\frac{a-1}{a}\sigma^2_{\tau\beta}$.

(i) the restrictions of covariance in (d) are dropped.

In the case of mixed model, the numerical calculations in the analysis of variance remain unchanged: that is the degree of freedom (d.f.), sum of squares (SS), and mean squares (MS) for each component are all calculated as in the cases of fixed and random effects models. However, to form the test statistics, we must examine the expected mean squares (EMS). They are shown in Table 1 [1]. In Table 1, we note that the difference between EMS of the restricted and unrestricted mixed models occurs only in the case of random factor B. Consequently, the hypothesis testing for the variance component of random factor B equal to zero, $\sigma_\beta^2 = 0$, is different. In the situation where experimenters are primarily interested in the fixed effect rather than the random effect, then they are modeling the random effect only to get a properly general treatment of fixed effect, the unrestricted version should be preferred to use because it is easier to explain since it does not depend on the assumptions in (b) and (d). Another argument for using the unrestricted version is that in the unbalanced case, it is very difficult to use the restricted version [5]. It can be seen that, the unrestricted version is more flexible than restricted version in that it can be used in both balanced and unbalanced designs. But while in the design which has limited resource in each level of the random factor B is likely to produce a negatively correlated interaction for different A-levels of the same B-level, so the restricted version will reflect this [4].

**Table 1** Expected mean squares and test statistics for two-way restricted and unrestricted mixed models.

| Source of variation | Restricted mixed model | | Unrestricted mixed model | |
|---|---|---|---|---|
| | EMS | Null Hypothesis; Test statistics | EMS | Null Hypothesis; Test statistics |
| A | $\sigma^2 + n\sigma_{\tau\beta}^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$ | $H_o: \tau_i = 0; F = \frac{MSA}{MSAB}$ | $\sigma^2 + n\sigma_{\tau\beta}^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$ | $H_o: \tau_i = 0;$ $F = \frac{MSA}{MSAB}$ |
| B | $\sigma^2 + an\sigma_\beta^2$ | $H_o: \sigma_\beta^2 = 0; F = \frac{MSB}{MSE}$ | $\sigma^2 + n\sigma_{\tau\beta}^2 + an\sigma_\beta^2$ | $H_o: \sigma_\beta^2 = 0;$ $F = \frac{MSB}{MSAB}$ |
| AB | $\sigma^2 + n\sigma_{\tau\beta}^2$ | $H_o: \sigma_{\tau\beta}^2 = 0;$ $F = \frac{MSAB}{MSE}$ | $\sigma^2 + n\sigma_{\tau\beta}^2$ | $H_o: \sigma_{\tau\beta}^2 = 0$ ; $F = \frac{MSAB}{MSE}$ |
| Error | $\sigma^2$ | - | $\sigma^2$ | - |

**Restricted and Unrestricted Models in Balanced Two-Way Mixed Model Using SAS and Minitab**

Nowadays, many of the well-known statistical softwares which can analyze data from the experimental design such as SAS, Minitab, SPSS, and R exist. In this review article, we presented only the instructions to use the programs SAS version 9.2 and Minitab version 15 and also presented the results of comparing the two programs. In SAS, the procedures that can handle the balanced mixed model are PROC GLM and PROC MIXED. It is noteworthy that PROC MIXED is recommended for analyzing the mixed model to avoid the pitfalls of PROC GLM [6]. In the GLM procedure assume that all effects are fixed, so that all tests and estimability checks for this procedure is based on a fixed effects model. This is true even if a RANDOM statement is used. SAS basically uses only unrestricted models in both procedures [6-9]. However, in the GLM procedure, users still have the flexibility to perform hypothesis testing for each effect specified in the model, using appropriate error terms by H and E option of TEST statement to specify numerator and denominator effects, respectively. But be aware of that this statement does not check any of the assumptions underlying the F statistic; therefore to help validate a test, the TEST option of the RANDOM statement should be used [8]. The code of SAS and the commands of Minitab to deal with the balanced mixed model are given in the below example.

**Pearl Data Example**

The data set considered in this review article is obtained from Neter, et al. [10]. They consider the production of limitation pearls and interested to study the effects of two factors on the market value of the pearls (dependent variable: y). The main factor of interest (factor A) is the number of coats of a special lacquer applied to the opalescent plastic bead used as the base of the pearl. Factor A has three levels (6, 8, and 10 coats) that were fixed in advance. The other factor (factor B) was the batch of the beads used in the study. Each batch consisted of 12 beads equally divided among the 3 levels of factor A. There were a total of 4 batches which can be regarded as a random sample of batches from the bead production process. To apply the linear statistical model in (1), we have factor A as fixed with only three levels; factor B as random with four levels from the population of all levels and the replications within each cell are equal to four units. In this example, we separate the cases of study into two cases. The first case, suppose that there exist a limited resources such as materials and human resources in each batch, cause the negatively relationship between the final market value of pearls in each batch may occur, so the suitable model is restricted. The second case, suppose the problem of limited resources does not occur then the proper model should be unrestricted.

The SAS's code and all necessary outputs from PROC GLM and PROC MIXED for both restricted and unrestricted cases [8-9] are displayed in Figure 1-4.

Thaksin.J., Vol.14 (1) January - June 2011

```
data pearl;
 do A = 1 to 3;
  do B = 1 to 4;
   do rep = 1 to 4;
    input y @@;
    output;
   end;
  end;
 end;
cards;
72.0  74.6  67.4  72.8  72.1  76.9  74.8  73.3
75.2  73.8  75.7  77.8  70.4  68.1  72.4  72.4
76.9  78.1  72.9  74.2  80.3  79.3  76.6  77.2
80.2  76.6  77.3  79.9  74.3  77.6  74.4  72.9
76.3  74.1  77.1  75.0  80.9  73.7  78.6  80.2
79.2  78.0  77.6  81.2  71.6  77.7  75.2  74.4
;
title 'Unrestricted Mixed Model';
proc glm data = pearl;
 class A B;            /*classification variables to be used in the model*/
 model y = A|B;        /*Specify dependent variable = all fixed and random effects*/
 random B A*B / test; /*Specify all random effects including random interactions*/
                       /*Test option to performs hypothesis tests for each effect specified
                         in the model, using appropriate error terms as determined by EMS*/
run;
proc mixed data = pearl method = type3;
              /*This example type3 (Method of Moments - ANOVA Table, Type III SS)
                was specified to conform with GLM*/
 class A B;     /*classification variables to be used in the analysis*/
 model y = A;   /*Specify dependent variable = fixed effects*/
 random B A*B; /*Specify all random effects including random interactions*/
run;
title 'Restricted Mixed Model';
proc glm data = pearl;
 class A B;           /*classification variables to be used in the model*/
 model y = A|B;       /*Specify dependent variable = all fixed and random effects*/
 random B A*B;        /*Specify all random effects including random interactions*/
 test H = A E = A*B; /*To request additional F tests that use other effects as error terms
                        H = effects which are used as hypothesis (numerator) effects
                        E = effect which is use as the error (denominator) term*/
run; quit;
```

**Figure 1** SAS's code of pearl data example for restricted and unrestricted cases.

```
                              The GLM Procedure

          Source                 Type III Expected Mean Square

          A                      Var(Error) + 4 Var(A*B) + Q(A)

          B                      Var(Error) + 4 Var(A*B) + 12 Var(B)

          A*B                    Var(Error) + 4 Var(A*B)

                   Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: y

       Source              DF     Type III SS     Mean Square    F Value    Pr > F

       A                    2      150.387917       75.193958     243.60    <.0001
       B                    3      152.851667       50.950556     165.06    <.0001

       Error: MS(A*B)       6        1.852083        0.308681

       Source              DF     Type III SS     Mean Square    F Value    Pr > F

       A*B                  6        1.852083        0.308681       0.06    0.9988

       Error: MS(Error)    36      173.625000        4.822917
```

**Figure 2** SAS's output of pearl data example by PROC GLM for unrestricted case.

**The Mixed Procedure**

Type 3 Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | Expected Mean Square | Error Term |
|---|---|---|---|---|---|
| A | 2 | 150.387917 | 75.193958 | Var(Residual) + 4 Var(A*B) + Q(A) | MS(A*B) |
| B | 3 | 152.851667 | 50.950556 | Var(Residual) + 4 Var(A*B) + 12 Var(B) | MS(A*B) |
| A*B | 6 | 1.852083 | 0.308681 | Var(Residual) + 4 Var(A*B) | MS(Residual) |
| Residual | 36 | 173.625000 | 4.822917 | Var(Residual) | . |

Type 3 Analysis of Variance

| Source | Error DF | F Value | Pr > F |
|---|---|---|---|
| A | 6 | 243.60 | <.0001 |
| B | 6 | 165.06 | <.0001 |
| A*B | 36 | 0.06 | 0.9988 |
| Residual | . | . | . |

**Figure 3** SAS's output of pearl data example by PROC MIXED for unrestricted case.

**The GLM Procedure**

Dependent Variable: y

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| A | 2 | 150.3879167 | 75.1939583 | 15.59 | <.0001 |
| B | 3 | 152.8516667 | 50.9505556 | 10.56 | <.0001 |
| A*B | 6 | 1.8520833 | 0.3086806 | 0.06 | 0.9988 |

Dependent Variable: y

Tests of Hypotheses Using the Type III MS for A*B as an Error Term

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| A | 2 | 150.3879167 | 75.1939583 | 243.60 | <.0001 |

**Figure 4** SAS's output of pearl data example by PROC GLM for restricted case.

For Minitab, the commands in menu bar to analyze this data are Stat --> ANOVA --> Balanced ANOVA, which the dialog box is shown in Figure 5.

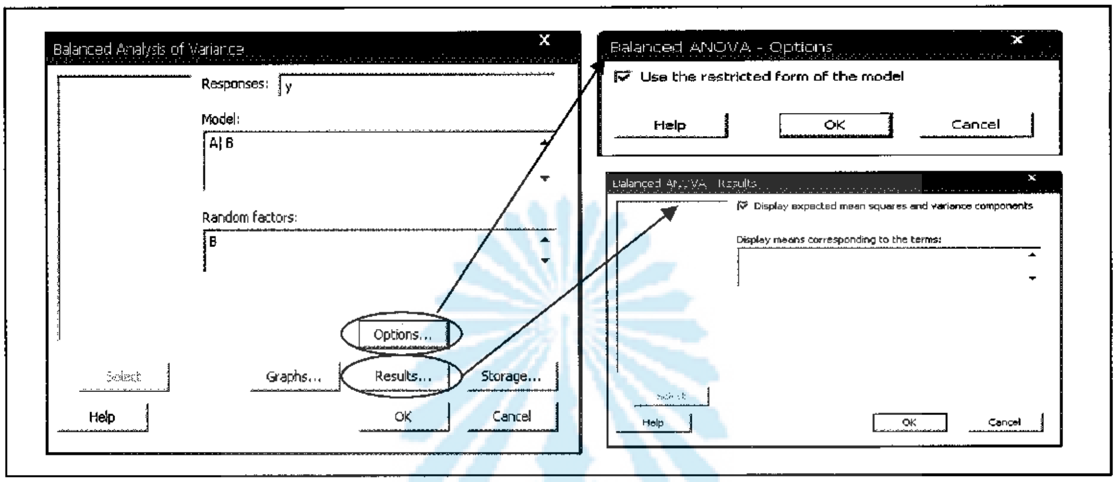


**Figure 5** Minitab's commands of pearl data example for restricted case.

Specify the dependent variable (y) in the box "Responses:". There are two ways to specify the model in the box "Model:" which give the similar result, that are either specify A B A*B or put a vertical bar between the designations of the two main effects: A | B. The random effect B must be declared in "Random factors:" The EMS is obtained from selecting "Results…" and then select "Display expected mean squares and variance components". The default of mixed model in Minitab is an unrestricted model. For the restricted model, click "Options…" and select "Use the restricted form of the model". The Minitab's outputs for restricted and unrestricted models are shown in Figure 6-7, respectively.

**ANOVA: y versus A, B**

```
Factor  Type    Levels  Values
A       fixed        3  1, 2, 3
B       random       4  1, 2, 3, 4

Analysis of Variance for y

Source  DF       SS      MS       F      P
A        2  150.388  75.194  243.60  0.000
B        3  152.852  50.951   10.56  0.000
A*B      6    1.852   0.309    0.06  0.999
Error   36  173.625   4.823
Total   47  478.717

S = 2.19611   R-Sq = 63.73%   R-Sq(adj) = 52.65%

                                   Expected Mean Square
              Variance   Error     for Each Term (using
   Source    component   term      restricted model)
1  A                        3      (4) + 4 (3) + 16 Q[1]
2  B             3.844      4      (4) + 12 (2)
3  A*B          -1.129      4      (4) + 4 (3)
4  Error         4.823             (4)
```

**Figure 6** Minitab's output of pearl data example for restricted case.

```
ANOVA: y versus A, B

Factor  Type    Levels  Values
A       fixed        3  1, 2, 3
B       random       4  1, 2, 3, 4

Analysis of Variance for y

Source  DF       SS      MS       F      P
A        2  150.388  75.194  243.60  0.000
B        3  152.852  50.951  165.06  0.000
A*B      6    1.852   0.309    0.06  0.999
Error   36  173.625   4.823
Total   47  478.717

S = 2.19611   R-Sq = 63.73%   R-Sq(adj) = 52.65%

                                 Expected Mean Square
             Variance  Error     for Each Term (using
   Source   component   term     unrestricted model)
1  A                       3     (4) + 4 (3) + Q[1]
2  B            4.220      3     (4) + 4 (3) + 12 (2)
3  A*B         -1.129      4     (4) + 4 (3)
4  Error        4.823            (4)
```

**Figure 7** Minitab's output of pearl data example for unrestricted case.

In this example, the results of both restricted and unrestricted models are similarly conclusions. There is insignificance of the interaction between coat (A) and batch (B) whereas all two main effects are significant. It should be noted that, the restricted and unrestricted models give the difference of F values for random factor B, i.e., 10.56 and 165.06, respectively. Since, both models give the high values of F statistics hence the decision of both is similarly, that is the null hypothesis $\sigma_\beta^2 = 0$ is rejected.

## Conclusions

The restricted and unrestricted models are two versions of model in mixed model theory. The unrestricted model is more suitable in many situations, such as in the unbalanced design or in the case of experimenters primarily interested to find a proper level in the fixed effect rather than in the random effect. While, the restricted model is appropriate for the design which has limited resource in each level of the random factor B and then influence to produce the negatively correlated interaction for different levels of fixed factor A at the same level of random factor B. The F values for the random effect of restricted and unrestricted models are different. Meanwhile, the fixed and interaction effects of both models have the same F values. This difference of F values that occurs may result in the decisions of experiment are not the same.

Generally, SAS is a program that is more flexible than Minitab. Sometimes there are things that we can do in SAS but cannot do in Minitab [5]. Minitab is a small program which may be suitable for studying in the classroom. It is surprising, particularly in the case of the restricted model; SAS can't be done with the program itself. But, it will require the knowledge of researchers about EMS. This means that, if researchers do not have the knowledge enough about statistics, especially for finding the EMS, and want to use "restricted" two- or multi-way mixed model, the researcher should not use SAS because it gives the F values of the "unrestricted" case. While, the program Minitab will not be affected of this problem.

**Acknowledgements**

This review article is (partially) supported by the Centre of Excellence in Mathematics, the Commission of Higher Education, Thailand. I would like to give special thanks to Assoc. Prof. Jirawan Jitthavech, National Institute of Development Administration (NIDA), for her comments and suggestions. The author is also thankful for the help and support of several people in my organization, Thaksin University.

**References**

[1] Montgomery, D.C. (2006). **Design and Analysis of Experiments.** 6th ed. New York: Wiley.

[2] Doncaster, C.P. and Davey, A.J.H. (2007). **Analysis of Variance and Covariance: How to Choose and Construct Models for the Life Sciences.** Cambridge: Cambridge University Press.

[3] Driscoll, M.F. and Borror, C.M. (2000). Sum of Squares and Expected Mean Squares in SAS. **Quality and Reliability Engineering International.** 16 (5), 423-433.

[4] Smith, M.K. (2011, February 24). **Mixed Models (Sections 17.7 – 17.8).** Retrieved on February 24, 2011 from http://www.ma.utexas.edu/users/mks/384E09/mixedmodslides.pdf

[5] Arnold, S. (2011, February 24) **Lesson 7: Random and Mixed Models.** Retrieved on February 24, 2011 from http://www.stat.psu.edu/online/courses/stat502/07_mixed/06_mixed_mixed.html

[6] Cheng, J., Olbricht, G., Gunaratna, N., Kendall, R., Lipka, A., Paul, S. and Tyner, B. (2011, February 24). **Mixed Models.** Retrieved on February 24, 2011 from www.stat.purdue.edu/~bacraig/SCS/mixed.pdf

[7] Wang, J.C. (2011, February 23). **About Expected Mean Squares.** Retrieved on February 23, 2011. From http://www.stat.wmich.edu/wang/664/notes/ems.pdf

[8] SAS/STAT® 9.2. (2011, February 23). User's Guide: **The GLM Procedure (Book Excerpt).** Retrieved on February 23, 2011 from http://support.sas.com/documentation/cdl/en/statugglm/61789/PDF/default/statugglm.pdf

[9] SAS/STAT® 9.2. (2011, March 3). User's Guide: **The Mixed Procedure (Book Excerpt).** Retrieved on March 3, 2011 from http://support.sas.com/documentation/cdl/en/statugmixed/61807/PDF/default/statugmixed.pdf

[10] Neter, J., Wasserman, W. and Kutner, M.H. (2006). **Applied Linear Statistical Models.** 4th ed. Irwin: McGraw-Hill.