

## SD14

# All Possible Mixed Model Selection - A User-friendly SAS Macro Application

George C J Fernandez, University of Nevada - Reno, Reno NV 89557

### ABSTRACT

A user-friendly SAS macro application to perform all possible model selection of fixed effects including quadratic and cross products in the presence of random and repeated measures effects using SAS PROC MIXED is available. This macro application will complement the model selection option currently available in the SAS PROC REG for multiple linear regressions and the experimental SAS procedure GLMSELECT that focuses on the standard independently and identically distributed general linear model for univariate responses. Options are also included in this macro to select the best covariance structure associated with the user-specified fully saturated repeated measures model; to graphically explore and to detect statistical significance of user specified linear, quadratic, interaction terms for fixed effects; and to diagnose multicollinearity, via the VIF statistic for each continuous predictors involved in each model selection step. Two model selection criteria, AICC (corrected Akaike Information Criterion) and MDL (minimal description length) are used in all possible model selection and summaries of the best model selection are compared graphically. The differences in the degree of penalty factors associated with the model dimension between AICC and MDL are investigated. Complete mixed model analysis of final model including data exploration, influential diagnostics, and checking for model violations using the experimental ODS GRAPHICS option available in Version 9.13 is also implemented. Instructions for downloading and running this user-friendly macro application are included.

### INTRODUCTION

Model selection is usually carried out by the automated procedures built into the software including frequently used forward, backward, and stepwise model selection procedures. There is an extensive review and discussion on the theoretical aspects of model selection criteria and procedures (Burnham and Anderson 2002; Hoeting et.al 2006). All possible model selection of fixed effects in general linear model setup using delta AICC and delta SBC is implemented in the REGDIAG macro (Fernandez, 2000). However, all possible mixed model selection can be tedious, time consuming and complicated due to the presence of additional random terms and optimal variance-covariance structure associated with repeated measures (Littell et.al 2006; Hoeting et.al 2006; Kramer 2004). Ngo and Rand (2002) developed a SAS macro for performing mixed model selection for user-specified models. Keselman et .al (1998) proposed a SAS based method to select the best covariance structure in mixed model repeated measures analysis. However, to apply these SAS macros in model selection, SAS programming experience is a requirement. Kramer (2004) developed an automated model selection application using SAS Mixed and PERL codes for both fixed and random effects. Programming knowledge in PERL is required to use this application. This paper presents a practical solution for automated and efficient mixed model selection and model exploration using a user-friendly SAS macro application named ALLMIXED.

### MODEL SELECTION CRITERIA USED IN ALLMIXED MACRO

The general form of IC =  $-2 \log L + \text{Penalty factor (pf)}$

$-2 \log L$  is derived from PROC MIXED method = ML

$$\Delta -2 \log L = 2 \log L_j - 2 \log L_{min}$$

$-2 \log L_{ref} = -2 \log L$  derived from PROC MIXED method ML that contain optional random and repeated measure covariance parameter and user specified "Must-Have" fixed effects.

$$AIC = -2 \log L + 2(p+k+1) \quad (\text{Hoeting et. al 2006})$$

$$AIC\_C = -2 \log L + [2(p+k+1) (n/(n-p-k-2))] \quad (\text{Hoeting et. al 2006})$$

Where

$p$  = number of fixed effect terms  
 $k$  = number of random effect terms  
 $n$  = total sample size

\* In large sample AIC and AIC\_C are nearly equivalent

$$\Delta AIC\_C = AIC\_C_i - AIC\_C_{min}$$

$AICC_{sas}$  = AICC reported by SAS PROC Mixed using ML

$AICC_{REML}$  = AICC reported by SAS PROC Mixed using REML

$$MDL = 1/2 \{-2 \log L + [\log(n) (p+k+1)]\} \quad (\text{Hoeting et. al 2006})$$

$$\Delta MDL = MDL_i - MDL_{min}$$

$$BIC = -2 \log L + [\log(n) (p+k+1)] \quad (\text{SAS Institute 2006})$$

$$\text{Penalty factor \%} = (pf_i / -2 \log L_{ref}) * 100$$

## SAMPLE DATA

Simulated repeated measured data with 50 subjects and 5 repeated (Time) measures / subject.

- Covariance type = AR(1) with a correlation = 0.8
- Four levels of significant treatments are randomly assigned to 50 subjects
- Significant Covariates : X5, X5<sup>2</sup>, X15, X5\*X15
- Non-Significant Covariates: X1, X2, X3, X4, X6, X7, X8, X9, X10, X11, X12, X13, X14, X16, X17, X18, X19, X20

## ALL POSSIBLE MODEL SELECTION STEPS

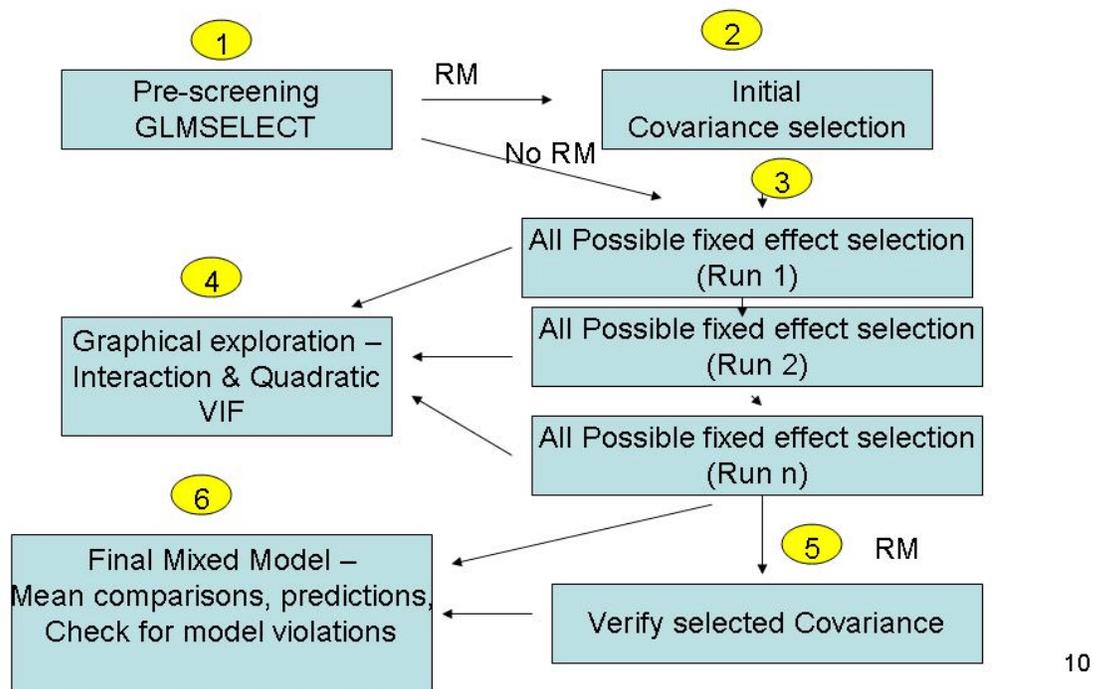
The recommended selection steps for performing the model selection in MIXED model is illustrated in Figure1. Although the recommended sequence of the steps are identified in the figure, it is not a requirement to follow the same sequence. Users are free to choose to run any model selection steps in any order the desire. However, before running these model selection steps the data must be prepared suitable for running the SAS PROC Mixed (Littell et.al 2006) procedure. The following types of PC data formats can be used with the ALLMIXED macro: SAS temporary and permanent data files, Microsoft excel or ACCESS data tables, COMMA or TAB delimited text file. Refer the ALLMIXED macro help file available from the author's website for more information regarding inputting data file name and its location in the macro-call window.

### STEP1: PRE-SCREENING

If the number of fixed effects exceeds 15, running all possible models will take very long time to complete. Therefore, under these circumstances, pre-screening is recommended to discard least contributing model terms. In pre-screening, the repeated measures covariance structure is ignored and the random effects could be treated as fixed. To discard the least important model terms and to select the user-specified number of effects, the LASSO (Tibshirani 1996) and LAR (Efron et al. 2004) methods implemented in the GLMSELECT (Cohen 2006) the experimental SAS procedure available in SAS version 9.13 is used.. For more information of the theory and selection

features refer SAS Institute (2006). Figures 2 and 3 show the summary table of 10 selected fixed effects from a total of 20 potential fixed effects contained in the simulated data based on LASS and LAR respectively. Both methods selected the same continuous (x5 x14 x15 x16 x17 ) and categorical contrasts ( TRT TIME) as important fixed effects in the pre-screening. These fixed effects will be used in the subsequent model selection steps. Refer the ALLMIXED macro help file available from the authors website for more information regarding inputting appropriate parameters in the macro-call window.

## Mixed model selection steps



**Figure 1** Recommended all possible model selection steps and their sequence

### STEP2: REPEATED MEASURES - INITIAL COVARIANCE TYPE SELECTION.

In a repeated measures modeling , the best covariance structure describing the correlation among the repeated measures should be identified first. The best covariance structure can be identified from different user-specified covariance structures by comparing the AICC statistic computed in PROC MIXED using REML method and select the covariance type which gives the smallest AICC value. In this example, four different covariance types, CS, AR(1), TOEP, and UN are compared in the full saturated model containing two categorical effects (TRT and TIME) and five continuous fixed effects selected from the previous pre-screening steps. The results of initial covariance type selection is graphically displayed in Figure4 and based on  $\Delta AICC_i (AICC_i - AICC_{min})$ , AR(1) can be identified as the best covariance type. Therefore, AR(1) covariance type will be used in the subsequent fixed effect selection. Refer the ALLMIXED macro help file available from the authors website for more information regarding inputting

appropriate parameters in the macro-call window.

**GLMSELECT : LASSO method of Tibshirani (1996) stop:10**  
**Data: simdata1 Resp: y Fixed effects: trt time x1-x20**

LASSO Selection Summary					
Step	Effect Entered	Effect Removed	Number Effects In	AICC	SBC
0	Intercept		1	5.8898	1223.9129
1	X15		2	5.8239	1210.9111
2	X5		3	5.8016	1208.7909
3	TRT_1		4	5.7484	1198.9428
4	X17		5	5.7131	1193.5370
5	TRT_3		6	5.7022	1194.2124
6	X14		7	5.6714	1189.9123
7	TIME_1		8	5.6710	1193.1654
8	X16		9	5.6520	1191.7699
9	TIME_5		10	5.6303	* 1189.6872 *
* Optimal Value Of Criterion					

Selection stopped at the first model containing the specified number of effects (10). 16

**Figure 2** Selected fixed effects including the continuous and categorical fixed effects based on LASSO selection method in SAS GLMSELECT

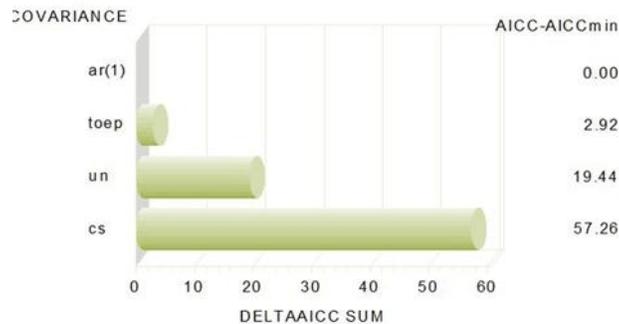
**GLMSELECT :Least angle regression (LAR) method of Efron et. al (2004).**  
**stop:10 Data: simdata1 Resp: y Fixed effects: trt time x1-x20**

LAR Selection Summary					
Step	Effect Entered	Effect Removed	Number Effects In	AICC	SBC
0	Intercept		1	5.8898	1223.9129
1	X15		2	5.8239	1210.9111
2	X5		3	5.8016	1208.7909
3	TRT_1		4	5.7484	1198.9428
4	X17		5	5.7131	1193.5370
5	TRT_3		6	5.7022	1194.2124
6	X14		7	5.6714	1189.9123
7	TIME_1		8	5.6710	1193.1654
8	X16		9	5.6520	1191.7699
9	TIME_5		10	5.6303	* 1189.6872 *
* Optimal Value Of Criterion					

Selection stopped at the first model containing the specified number of effects (10). 17

**Figure 3** Selected fixed effects including the continuous and categorical fixed effects based on LAR selection method in SAS GLMSELECT

**Covariances= cs ar(1) toep un**  
**Selecting RM covariance structure**



41

**Figure 4** Repeated Measure analysis covariance type selection based on smallest AICC.

**Step3: All possible model selection steps**

All combination of models associated with the user-specified fixed effects are generated by the ALLMIXED macro and their information criteria statistics, AIC\_C and MDL are compared in this step. Users can optionally differentiate certain fixed effects as “MUST HAVE” and other fixed effects as “SELECTABLE” in the all possible model selection. All combination of mixed model using the fixed effects listed in “SELECTABLE” category are generated in this step and the following statistics are estimated.

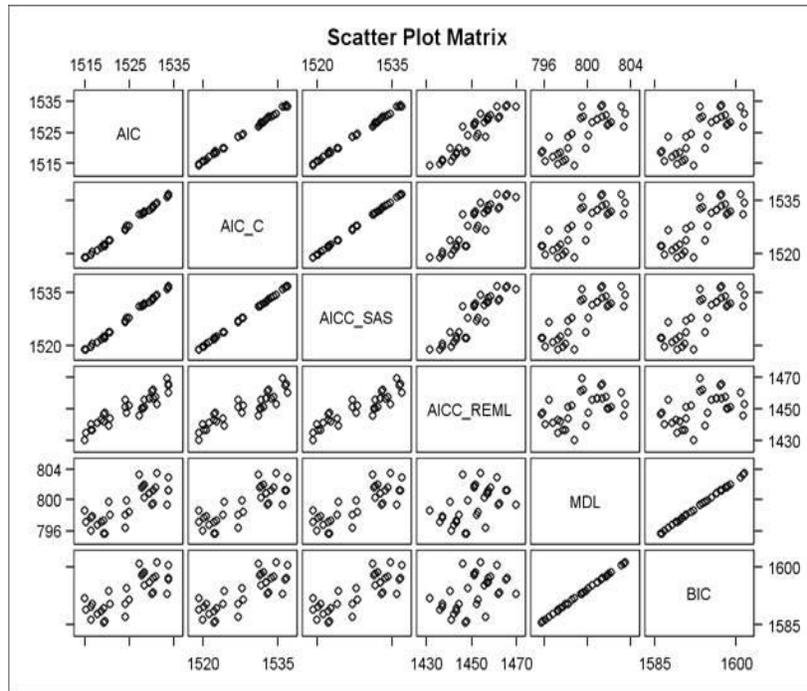
Variance inflation statistics (VIF) for each continuous predictor variables in the model.

Information criteria estimates based on REML:  $AICC_{reml}$

Information criteria estimates based on ML: AIC, AIC\_C,  $AICC_{sas}$ , MDL, and BIC. The relationships between AIC, AIC\_C,  $AICC_{sas}$ ,  $AICC_{reml}$ , MDL, and BIC are investigated by the rank correlations using the SAS PROC CORR and scatter plot matrix (Figure 5). Perfect rank correlations (1) were observed between (AIC\_C and  $AICC_{sas}$ ) and (MDL and BIC) indicating that these two sets of IC behave identically in the model selection. Furthermore AIC and AIC\_C behaved very similarly and the degree of penalty was similar when the number of fixed effects is relatively small ( $p=19$ ) compared with total number of observations ( $n=250$ ) in this simulated data. The rank correlations between  $AICC_{reml}$  and AIC statistics computed by ML method (AIC, AIC\_C,  $AICC_{sas}$ ) were not perfectly correlated and  $AICC_{reml}$  behave differently from ML based IC. Big differences were observed in the model selection performance of AIC based (AIC, AIC\_C, ) and BIC based (MDL) information criteria as evident by the rank correlation.

All IC statistics reported here are made out of two components: a) Log likelihood estimate ( $-2 \log L$ ) b) penalty factor ( $pf$ ). For a given model,  $\log L$  value is constant and influenced by degree of model fit, variable included and not included in the model, presence of influential outliers, model specification errors, and severe multicollinearity. The penalty factor is made out of number of fixed ( $p$ ) and random effects ( $k$ ) and the sample size ( $n$ ). When all possible

## Rank Correlation among IC statistics



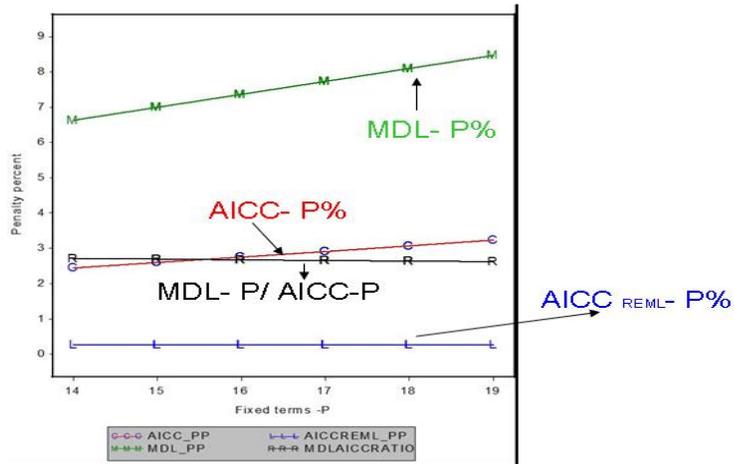
23

**Figure 5** Scatter plot matrix showing rank correlation among 4 AIC and 2 BC based Information criteria statistics in all possible mixed model selection.

model selection involving only the fixed effects are carried out, the sample size and the number of random effects become constant. Therefore, only the number of fixed factor becomes the determining component of the penalty factor. The relationship between penalty factor and the number of fixed effects between  $AIC_C$ ,  $AICC_{reml}$ , and MDL are shown in Figure 6. The penalty factor for the  $AICC_{reml}$  becomes constant because this penalty factor does not include any fixed effects and only the number of random effects (which is a constant) is included. The penalty factors for  $AIC_C$  and MDL shows a positive linear effect associated with the increase in the number of fixed effects. The MDL penalty factor % increased from 6.5% (Reference model) to about 8.5% whereas the  $AIC_C$  penalty factor % increased from 2.5% (Reference model) to about 3.2% when the number of fixed effects increased by 5 terms. Thus, the degree of penalty is about 3 times stronger for MDL than the  $AIC_C$  and this clearly evident in the ratio between MDL penalty % and  $AIC_C$  penalty and the relationship is clearly shown in Figure 6. These findings clearly confirm the earlier reports (Hoeting et. al 2006 ) that MDL statistic favors more parsimonious model than  $AIC_C$  based model selection.

The components of  $AIC_C$  and MDL are graphically compared in Figure 7. For a given model,  $\log L$  value is constant in the estimation of  $AIC_C$  and MDL and it decreases linearly with an increase in the number of fixed terms. But, within a subset (two, three, four variable subset), the  $\log L$  value varies a lot whereas all the models within a subset

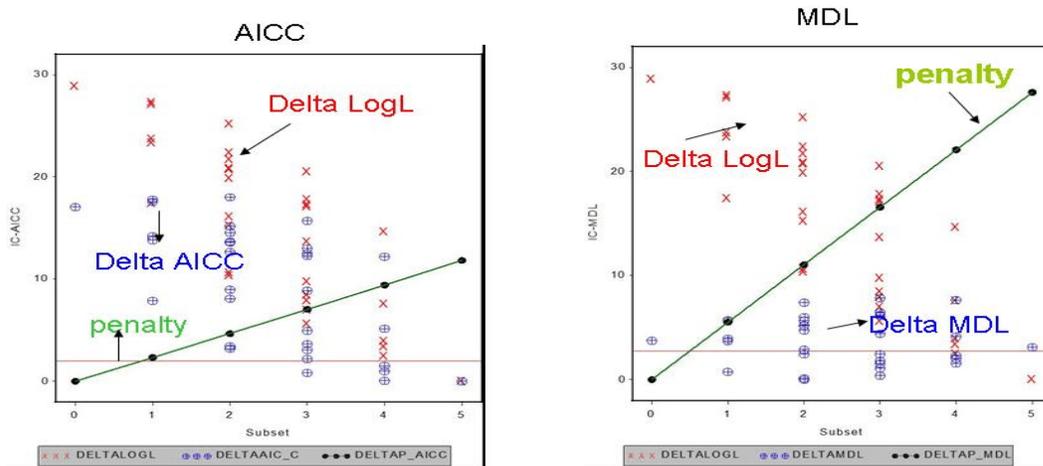
**Comparison of penalty (%) factors among IC statistics**



24

Figure 6 Comparison of penalty factor among AICC, AICC<sub>reml</sub> and MDL versus number of fixed effect terms.

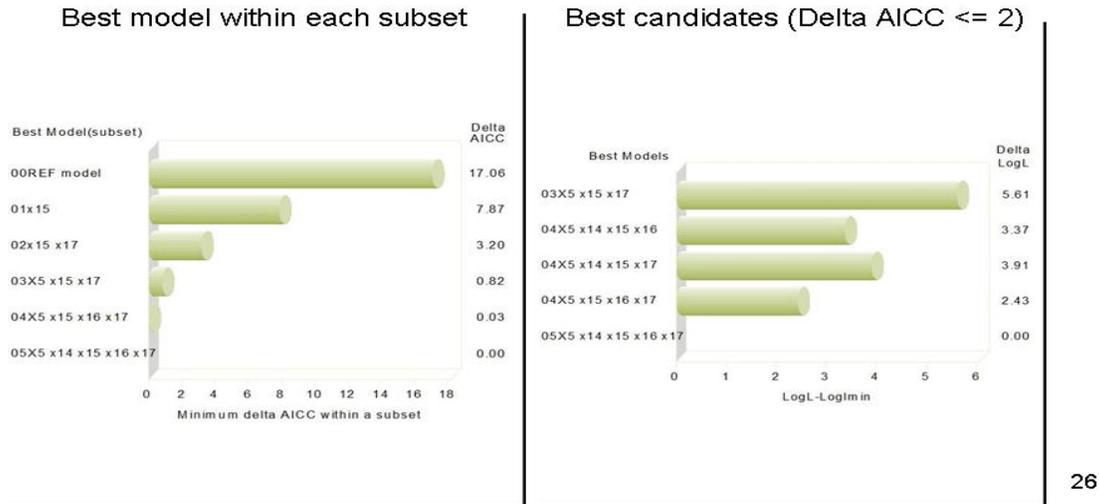
**Comparison of IC statistics**



25

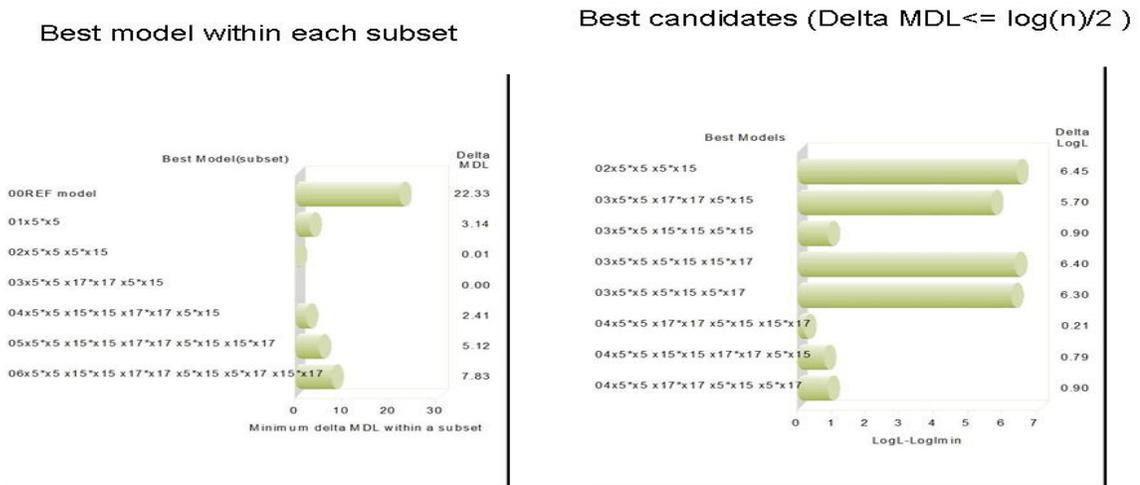
Figure 7 Comparison of the association between AICC and MDL components and the number of fixed effect terms

**Mixed model selection – delta AICC**



**Figure 8** Graphical display of best model within each subset and the list of best candidate models identified by the AICC.

**All possible Mixed model selection – MDL**



**Figure 9** Graphical display of best model within each subset and the list of best candidate models identified by the MDL

have the same penalty factor for both AIC\_C and MDL (Figure 7). Furthermore, MDL statistic favors parsimonious model (1,2 and 3 subsets) whereas AIC\_C statistic favors models with large number of model terms (3,4,5 subsets) especially in a large data set (50 subjects by 5 repeated measures) (Figure 7).

Graphical display of best models within each subset based on smallest  $\Delta AIC_C$  within each subset and the list of best candidate models ( $\Delta AIC_C \leq 2$ ) are shown in Figure 8. The full model with all five fixed effects was identified as the overall best model and 1, 3-variable model, 3, 4-variable models and 1, 5-variable full models were identified as the best candidate models based on  $\Delta AIC_C \leq 2$ . The  $\log L$  value always decreases with an increase in number of model terms. Graphical display of best models within each subset based on smallest  $\Delta MDL$  within each subset and the list of best candidate models ( $\Delta MDL \leq \log(n)/2$ ) are shown in Figure 9. The 3-variable model was identified as the overall best model and 1, 2-variable model, 4 3-variable models and 3, 4-variable full models were identified as the best candidate models based on ( $\Delta MDL \leq \log(n)/2$ ). The model selection results for this simulated large data (50 subjects, 5 repeated measures) clearly shows that  $\Delta MDL$  favors parsimonious model whereas  $\Delta AIC_C$  favors over estimated model. However, these findings needs to further verified by testing different types data sets with different sample sizes. Refer the ALLMIXED macro help file available from the authors website) for more information regarding inputting appropriate parameters in the macro-call window.

#### STEP4: GRAPHICAL EXPLORATION FOR MULTICOLLINEARITY AND MODEL SPECIFICATION ERROR

Severe multicollinearity (Variance inflation factor > 10) among predictor variables in mixed model analysis can result in unstable parameter estimates with inflated standard errors. When a fixed effect predictor involved in a collinear relationship is dropped from the model, the sign and size of the remaining predictor variable estimates can change dramatically. Therefore, presence of high degree of multicollinearity can impact fixed effect selection. Therefore, assessing the degree of multicollinearity for each of the continuous fixed effects in all possible model selection can help to select the best model from the set of best candidate models. Variable(s) not contributing multicollinearity could be preferred over the variables significantly contributing to multicollinearity. Figure 10 shows the box-plot display of VIF distribution for all the continuous predictors included in model selection. Because the data used in the study are simulated from known properties multicollinearity should not exists and it is clearly shown in Figure10 where VIF values were less than 2 for all the predictor variables. Also, to diagnose multicollinearity in each model selection step (when VIF value > 10) the VIF statistic for each continuous predictors involved in multicollinearity is sent to an output table for further exploration.

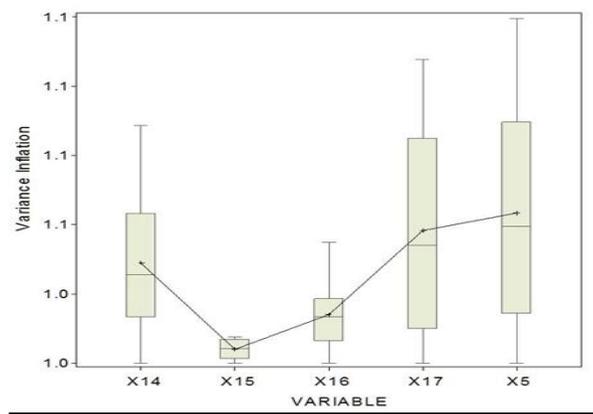
Model selection success can also be influenced by model specification error when significant higher order model terms (quadratic and cross-product) omitted from the mixed model. The need for an quadratic term or an interaction between any two predictor variables could be evaluated in the 'quadratic' or 'interaction detection plot respectively. To detect the need for a significant quadratic term, first fit the full model including the quadratic term for the given predictor variable ( $X_i$ ) and examine the Type III P-value for statistical significance and output the predicted values ( $YHAT_{full}$ ) for the full model. Then drop both the linear and the quadratic terms for this given predictor from the model and estimate the predicted values for this reduced model ( $YHAT_{red}$ ). Then a graphical display between the  $\Delta yhat$  ( $YHAT_{full} - YHAT_{red}$ ) and  $X_i$  can reveal the nature and the strength of quadratic effects (Figure 11).

Similarly, to detect the need for a significant interaction term between two predictors, first fit the full model including the cross-product term for the two predictor variable ( $X_1$  and  $X_2$ ) and examine the Type III P-value for statistical significance of the interaction term and output the predicted values ( $YHAT_{full}$ ) for the full model. Then drop the cross product from the model and estimate the predicted values for this reduced model ( $YHAT_{red}$ ). Then a 3-D graphical display between the  $\Delta yhat$  ( $YHAT_{full} - YHAT_{red}$ ) and  $X_1$  and  $X_2$  can reveal the nature and the strength of interaction effects (Figure 12). Refer the ALLMIXED macro help file available from the authors website) for more information regarding inputting appropriate parameters in the macro-call window.

#### Step5: Final covariance type selection.(Optional step- for repeated measures model)

After several runs of all possible model selection steps, many data exploration, and multicollinearity checks, we can select the final fixed effect model. But, before finalizing the final mixed effect model it is important to verify whether the covariance type used in the model selection step is still the best type for the selected model. Again user-specified covariance types, can be compared and the final covariance type selection can be made based on  $\Delta AICC_j$  ( $AICC_j - AICC_{min}$ ) using PROC MIXED REML method. Refer the ALLMIXED macro help file available from the authors

**Box plot of VIF statistics by continuous predictor variables**

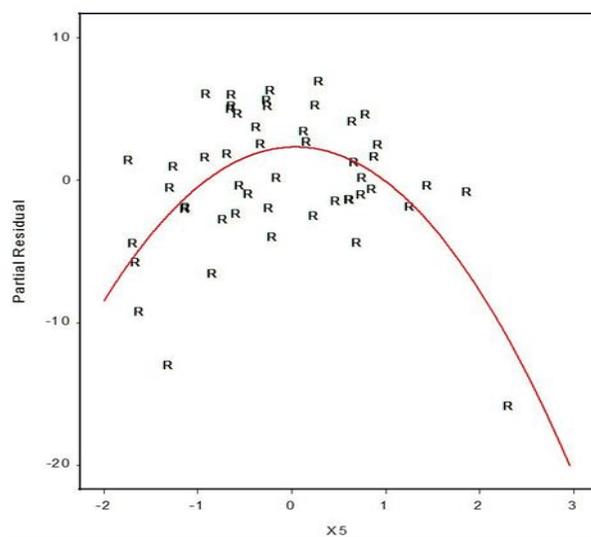


28

**Figure 10** Box-plot display of VIF statistic of all continuous predictor variables included in the possible model selection.

website) for more information regarding inputting appropriate parameters in the macro-call window.

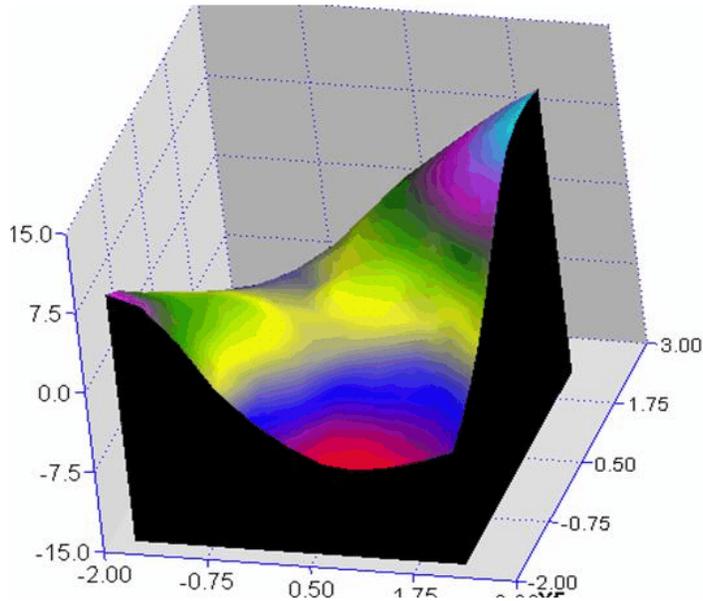
**Checking for quadratic effect –  $X5^2$  P-value:<0.0001**



36

**Figure 11** Graphical exploration and the statistical significance of the user-specified quadratic effect.

**Checking for interaction effect – X5\*X15 P-value:0.0004**



38

**Figure 12** Graphical exploration and the statistical significance of the user-specified cross product

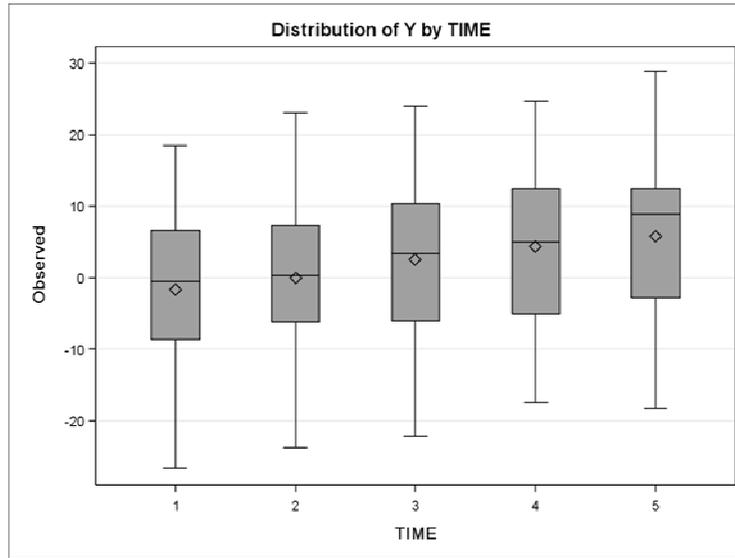
**Step6: Complete mixed model analysis**

After selecting the final repeated measures mixed model dimensions, complete mixed model analysis can be performed including the data exploration by box plots (Figure 13), mixed model analysis, LSMEAN comparisons with alphabet mean separation (Figure 14-15) suggested Saxton (2002), model predictions, checking for normality of studentized conditional residuals (Figure 16), and performing influential diagnostics (Figure 17) in one step. Refer the ALLMIXED macro help file available from the authors website for more information regarding inputting appropriate parameters in the macro-call window.

**AVAILABILITY OF THE ALLMIXED MACRO:**

Users can download the ALLMIXED .SAS macro-call file from the authors website at <http://www.ag.unr.edu/gf> and by clicking the "Running puppy dog" clip art. Save the ALLMIXED .SAS macro-call file in your PC first and open it in SAS display manager and submit to view the blue macro-call window (Figure 18) (You need to have access to INTERNET while running this SAS macro in your system). Input all the required macro input parameters and submit the macro to perform the all possible mixed model selection. Please refer the required SAS modules listed below for running this macro successfully.

Mixed model exploration – ODS GRAPHICS



37

Figure 13 Exploration of repeated measures-Time effect by box plot

Mixed model analysis data=simdata1 response=y  
 Mean separation based on [Saxton, A.M\(1998\) 23rd SUGI pp1243-1246](#)

Effect=TRT Method=Tukey-Kramer(P<.05) Table=1

TRT	TIME	LSmeans	Standard Error	Alpha	Lower CI	Upper CI	Alphabet Group
3	—	8.6556	1.3078	0.05	6.0344	11.2769	A
2	—	2.3088	1.4381	0.05	-0.5736	5.1913	B
1	—	-2.2621	1.0871	0.05	-4.4410	-0.08309	C

Effect=TIME Method=Tukey-Kramer(P<.05) Table=2

TRT	TIME	LSmeans	Standard Error	Alpha	Lower CI	Upper CI	Alphabet Group
—	5	6.6457	0.8735	0.05	4.9126	8.3789	A
—	4	5.1903	0.8735	0.05	3.4572	6.9235	A
—	3	3.4085	0.8735	0.05	1.6754	5.1416	B
—	2	0.6109	0.8735	0.05	-1.1222	2.3441	C
—	1	-1.3515	0.8735	0.05	-3.0846	0.3817	D

41

Figure 14 Main effect LSMEAN Comparison - using alphabet notation

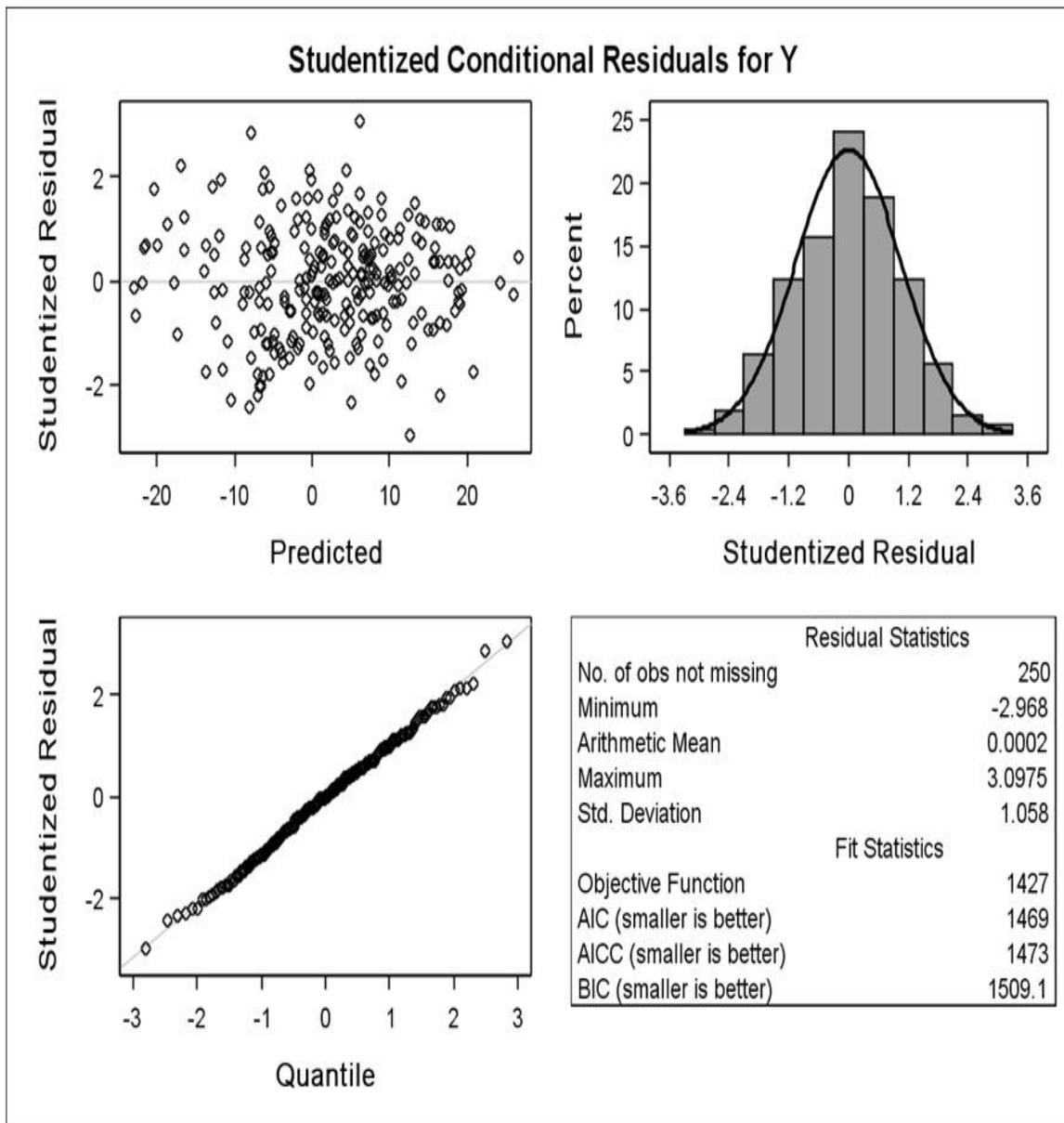
Mixed model analysis data=simdata1 response=y  
 Mean separation based on [Saxton, A.M\(1998\) 23rd SUGI pp1243-1246](#)

**Effect=TRT\*TIME** Method=Tukey-Kramer(P<.05) Table=3

TRT	TIME	LSmeans	Standard Error	Alpha	Lower CI	Upper CI	Alphabet Group
3	5	13.2799	1.5724	0.05	10.1595	16.4003	A
3	4	11.6835	1.5724	0.05	8.5631	14.8039	AB
3	3	10.0406	1.5724	0.05	6.9202	13.1610	ABC
3	2	6.6650	1.5724	0.05	3.5446	9.7854	CDE
2	5	5.6627	1.7168	0.05	2.2548	9.0706	ABCDEF
2	4	5.0490	1.7168	0.05	1.6411	8.4569	BCDEF
2	3	3.2151	1.7168	0.05	-0.1929	6.6230	CDEFGH
3	1	1.6093	1.5724	0.05	-1.5111	4.7297	FGHI
1	5	0.9946	1.3044	0.05	-1.5941	3.5834	DEFG
2	2	-0.6784	1.7168	0.05	-4.0863	2.7296	EGHI
1	4	-1.1615	1.3044	0.05	-3.7502	1.4273	FGHI
2	1	-1.7042	1.7168	0.05	-5.1121	1.7037	GH
1	3	-3.0302	1.3044	0.05	-5.6190	-0.4414	H
1	1	-3.9595	1.3044	0.05	-6.5482	-1.3707	GH
1	2	-4.1538	1.3044	0.05	-6.7426	-1.5651	H

42

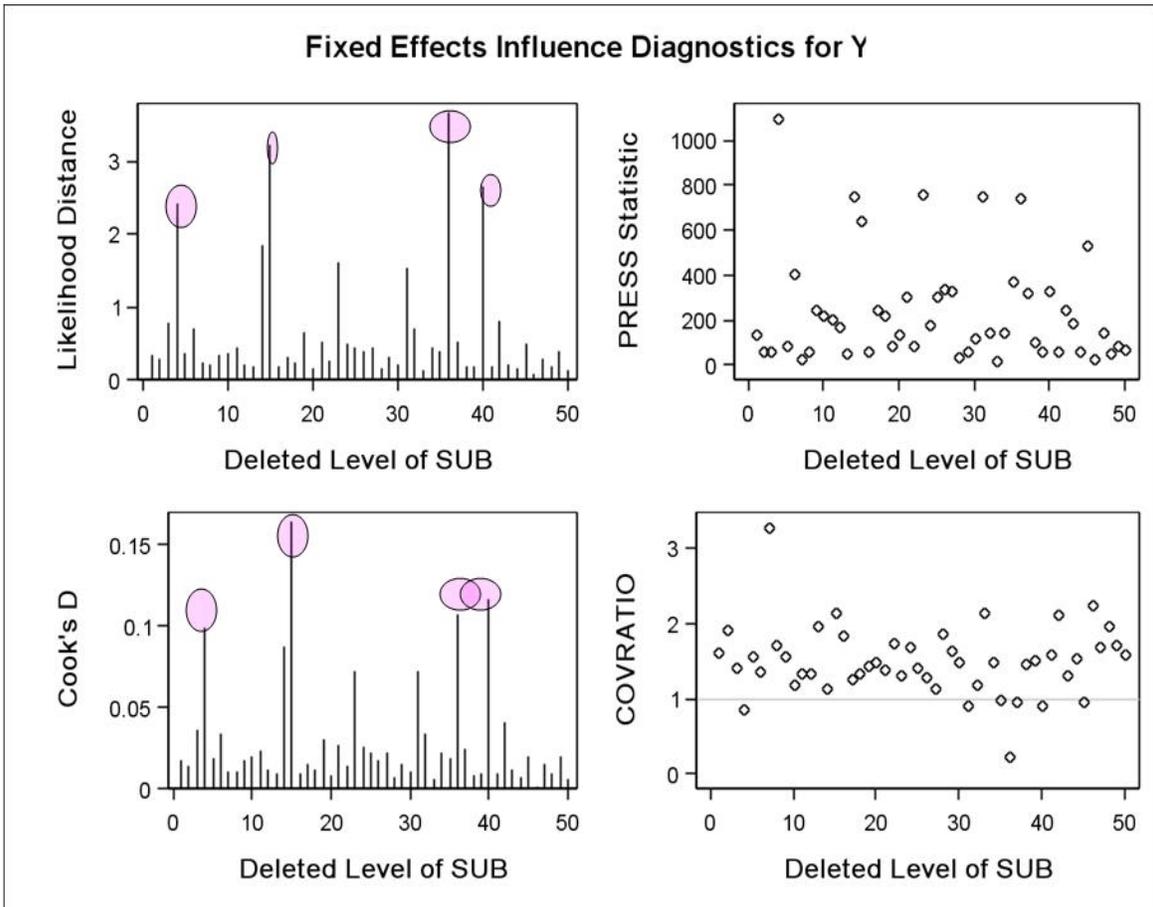
Figure 15 LSMEAN Comparison - Interaction means with alphabet notations



**Figure 16** Mixed model violation detection using studentized conditional residuals

**Required SAS Modules for Running the All mixed Sas Macro in Version 9.13:**

- SAS /STAT : PROCS MIXED, CORR, REG and GLMSELECT
- SAS/GRAPH: PROC GCHART, PROC GPLOT, PROC G3D
- SAS/BASE SAS ODS (RTF, HTML, PDF)
- SAS/ACCESS: PC FILES – PROC IMPORT and EXPORT



**Figure 17** Repeated measurers mixed model influential diagnostics - at the subject level

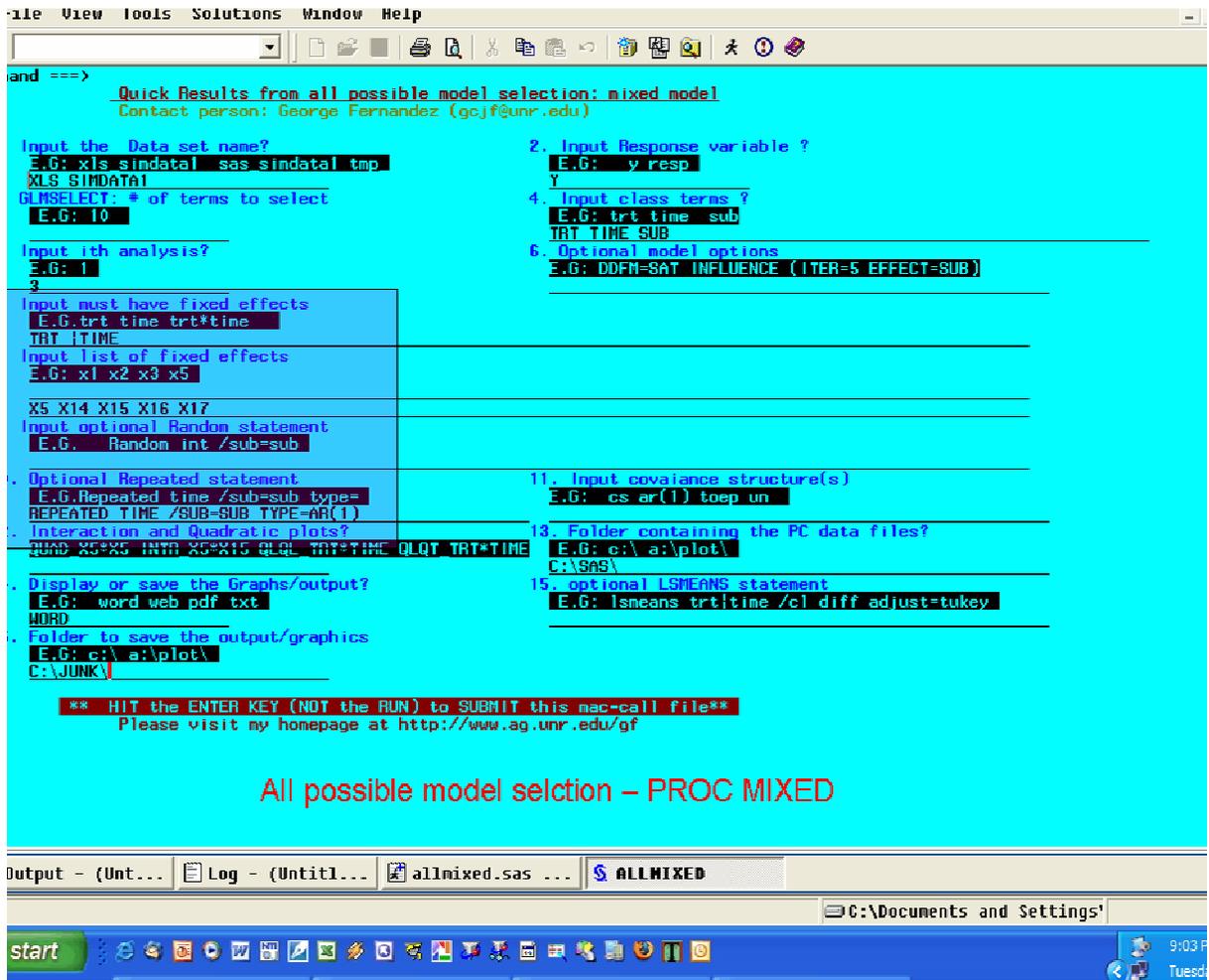


Figure 18 Screen shot of ALLMIXED macro-call window.

## SUMMARY

The main features of the user-friendly SAS macro application, ALLMIXED are summarized below:

- The users can input, temporary and permanent SAS data files, Microsoft Excel and Access and comma and TAB delimited text files as input data set.
- Users can input multiple response variable and perform all the model selection steps simultaneously.
- Users can optionally pre-screen the fixed effects and drop obvious non-significant fixed effects if the number of fixed effects exceed 15 using the SAS 9.1 experimental GLMSELECT procedure implemented within the macro. Two new model selection methods, LASSO and LAR are used in this macro to pre-screen the many fixed effects.
- In case of repeated measures mixed model analysis, the best covariance structure selection from the user specified covariance structures are implemented by comparing the AICC value estimated in the Proc Mixed using REML method and then best covariance structures is graphically identified by searching for the covariance structure with the smallest AICC value.
- Options for performing all possible fixed effect model selection with and without repeated and random effects and selecting the best candidate models using AICC and MDL estimates using PROC MIXED method ML. In this step, users can differentiate the “must-keep” effects and “selectable” effects. The all possible model selection will be performed using the fixed effects identified in the “Selectable” list of terms.
- Options are also available for graphical exploration and statistical significance of user specified linear, quadratic, interaction terms for fixed effects. Also, to diagnose multicollinearity (when VIF value > 10) the VIF statistic for each continuous predictors involved in each model selection step are sent to an output table. Also, a boxplot display of VIF estimates by all the continuous fixed effects are generated for the overall assessment of multicollinearity in the model selection process.
- Options are also available for performing complete mixed model analysis of final model including data exploration, influential diagnostics, and checking for model violations using the experimental ODS GRAPHICS option available in Version 9.1.
- Users can save all SAS output and graphics in Word, HTML, or PDF formats. In addition, full details all model selection diagnostic statistics are automatically sent to MS excel data tables. SAS log messages are automatically saved to external text log files and only the ERROR and WARNING messages are extracted and displayed as HTML output for easy error checks.
- Download instructions are given above to download this macro-call file and to perform all possible model selection.

## REFERENCES

- Burnham, K. P., and Anderson D. R. (2002) Model selection and inference: a practical information theoretic approach. (Second edition) Springer-Verlag, New York, New York, USA.
- Cohen R. A (2006) Introducing the GLMSELECT PROCEDURE for Model Selection SUGI31 proceedings <http://www2.sas.com/proceedings/sugi31/207-31.pdf>
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), “Least Angle Regression (with discussion), Annals of Statistics, 32, 407–499.
- Fernandez, G (2002) Data Mining using SAS applications Book 384 p CRC-Chapman Hall FL USA
- Hoeting, J.A, Davis R.A Merton A.A and Thompson S. E (2006) Model selection for Geostatistical Models Ecological Applications, 16(1), pp. 87–98
- Keselman, H. J., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1998) A comparison of two approaches for selecting covariance structures in the analysis of repeated measurement, 27, Communications in Statistics, Simulation & Computation 591-604.
- Littell, R.C, Milliken, G A., Stroup, WW, and Wolfinger, R D. (2006). SAS System for Mixed Models (second edition), Cary, NC: SAS Institute Inc.
- Kramer M (2004) Automatic model selection in the mixed model framework KSU applied statistics conference proceedings
- Ngo, L and Rand, R. (2002). Model Selection in Linear Mixed Effects Models Using SAS® Proc Mixed. SUGI 22 <http://www2.sas.com/proceedings/sugi22/STATS/PAPER284.PDF>
- SAS Institute (2006) The GLMSELECT procedure (Experimental) Cary, NC <http://support.sas.com/rnd/app/papers/glmselect.pdf>
- Saxton, A.M A Macro for Converting Mean Separation Output to Letter Groupings in PROC MIXED SUGI 23 <http://www2.sas.com/proceedings/sugi23/Stats/p230.pdf>
- Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso Journal of the Royal Statistical Society

Series B, 58, 267–288.

## **CONTACT INFORMATION**

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

Name: Dr. George C. Fernandez,  
Enterprise: University of Nevada - Reno  
Address: CABNR/204 Reno, NV 89557  
Work phone: (775)-784-4206  
Email: [gcjf@unr.edu](mailto:gcjf@unr.edu)  
Web: [Http://www.ag.unr.edu/gf](http://www.ag.unr.edu/gf)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

