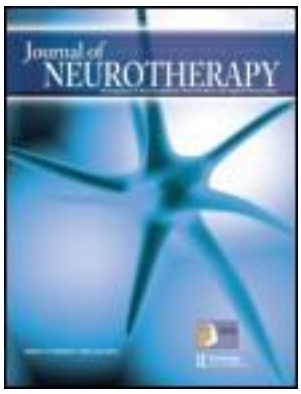

Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience

Mixed General Linear Model Analysis of Quantitative Electroencephalographic (qEEG) Data

Donald R. Bars PhD [a] & Christian Schindler PhD [b]

[a] Kantonsspital Basel Universitatskinien

[b] Institute of Social and Preventive Medicine, University of Basel , Switzerland

Published online: 08 Sep 2008.

To cite this article: Donald R. Bars PhD & Christian Schindler PhD (2004) Mixed General Linear Model Analysis of Quantitative Electroencephalographic (qEEG) Data, Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience, 8:4, 63-75

To link to this article: http://dx.doi.org/10.1300/J184v08n04_05

PLEASE SCROLL DOWN FOR ARTICLE

© International Society for Neurofeedback and Research (ISNR), all rights reserved. This article (the "Article") may be accessed online from ISNR at no charge. The Article may be viewed online, stored in electronic or physical form, or archived for research, teaching, and private study purposes. The Article may be archived in public libraries or university libraries at the direction of said public library or university library. Any other reproduction of the Article for redistribution, sale, resale, loan, sublicensing, systematic supply, or other distribution, including both physical and electronic reproduction for such purposes, is expressly forbidden. Preparing or reproducing derivative works of this article is expressly forbidden. ISNR makes no representation or warranty as to the accuracy or completeness of any content in the Article. From 1995 to 2013 the *Journal of Neurotherapy* was the official publication of ISNR (www. Isnr.org); on April 27, 2016 ISNR acquired the journal from Taylor & Francis Group, LLC. In 2014, ISNR established its official open-access journal *NeuroRegulation* (ISSN: 2373-0587; www.neuroregulation.org).

THIS OPEN-ACCESS CONTENT MADE POSSIBLE BY THESE GENEROUS SPONSORS

TECHNICAL NOTES

The purpose of the "Technical Notes" section is to provide detailed technical descriptions and illustrations of software, hardware, and techniques within our technically sophisticated field. This section does not have the depth of review that scientific research articles require, though submissions are reviewed for technical accuracy. This column will also contain authors' opinions and value judgments. The personal opinions expressed in the column are exactly that, the opinions and experiences belonging to the individual authors.

Mixed General Linear Model Analysis of Quantitative Electroencephalographic (qEEG) Data

Donald R. Bars, PhD
Christian Schindler, PhD

ABSTRACT. This paper describes a mixed general linear analysis of the quantitative electroencephalogram (qEEG). The modeling is similar to regression, which builds a regression or 'best-fit' model for the data

Donald R. Bars, a private investigator in the field of neurophysiology, recently moved to Ames, Iowa from Basel, Switzerland where he was consulting with the Kantonsspital Basel Universitatskinien. Christian Schindler works for the Institute of Social and Preventive Medicine, University of Basel, Switzerland.

Address correspondence to: Donald R. Bars (Email: dbars2001@yahoo.com).

Journal of Neurotherapy, Vol. 8(4) 2004
Copyright © 2004 ISNR. All rights reserved.
Digital Object Identifier: 10.1300/J184v08n04_05

structure but, in addition, provides for correlations between observations. A mixed linear model states that data consists of two parts: *fixed effects* and *random effects*. Fixed effects determine the expected values of the observations, while random effects account for the stochastic deviations from these expected values both between and within individuals. Since errors are independent between subjects, the deviations from the expected values may also be modeled using a repeated measures approach. The term 'repeated measures' in this model refers to data with multiple observations from one specific source. It is reasonable to assume that these observations from the same source are correlated, even if only slightly, in some measurable way. Consequently, statistical analysis of repeated measures data gives a more accurate prediction capability when the issue of covariation between these measures is addressed.

With mixed model methodology now available (e.g., the mixed procedure [Mixed PROC] of the SAS® system), the covariance structure can be incorporated into the statistical model. Disregarding potential random effects not specific to single individuals and absorbing potential within-subject random effects into the covariance matrix allows one to work with a simplified model.

The use of a mixed procedure and its method of modeling the data structure appear to provide an accurate and objective method of analysis resulting in quantifiable equations for testing predictions. Essentially, this method allows the physiological pattern of each individual in the study, not related to any other variable, to be represented and accounted for in the model. Several comparative examples will be used to highlight the information that can be hidden in data structures depending on the type of statistical analysis used.

Copyright © 2004 ISNR. All rights reserved.

KEYWORDS. General linear model (GLM), covariance, Fourier, quantitative electroencephalography

This paper describes the application of a statistical method for the analysis of the quantitative information created by digitizing an electroencephalogram (qEEG). The modeling is similar to regression, which builds a regression or 'best-fit' model for the data structure, but, in addition, provides for correlations between observations. A linear mixed model states that data consists of two parts, *fixed effects* and *random effects*. The fixed effects determine the expected values of the observa-

tions (e.g., the amount of activity in a specific brain frequency band for a specific electrode location) while random effects account for random deviations from these expected values both between and within individuals. If errors are independent between subjects, the deviations from the expected values may also be modeled using a repeated measures approach (i.e., by modeling the covariance structure of residuals). The term 'repeated measures' in this model refers to data with multiple observations from one specific source (i.e., the data gathered simultaneously from a collection of electrode sites on one individual). It is reasonable to assume that these observations from the same individual are correlated, even if only slightly, in some measurable way. Consequently, statistical analysis of repeated measures data gives a more accurate prediction capability when the issue of covariation between these measures is addressed.

With advances in technology and the mixed model methodology now available (e.g., the Mixed Procedure [MIXED PROC] of the SAS® system; SAS OnlineDoc™, Version 8), the covariance structure can more easily be incorporated into the statistical model (Cnaan, Laird, & Slasor, 1997; Koch, Tangen, Jung, & Amara, 1998; Littell, Pendergast, & Natarajan, 2000). As Littell et al. suggested in 2000, "Ignoring covariance structure may result in erroneous inference, and avoiding it may result in inefficient inference." The examples presented in this paper will highlight what can happen when not accounting for individual covariance. Hopefully, by the time you finish this paper you will understand the model and be able to utilize the method, primarily by a simple change in variable names, with any clinical questions that arise in your practice.

In typical comparative qEEG experiments subjects are assigned to different classes with the dependent variable being the subject's mean (y1) of data across k different electrode sites and 1 different frequency band. Mathematically speaking, the variable y1 is a vector with $k \times 1$ elements per subject. Usually the most important independent variable modeled concerns differences between groups though many times there are multiple independent variables that must be considered. Random effects result from variation between the subjects in the groups modeled by repeated measures of data recorded from different electrode sites and evaluated in frequencies. While repeated measures can be accounted for in the general linear model (GLM), the covariance of the data within each individual is not modeled.

With the mixed procedure, the fact that measures from the same person at different electrodes are correlated is taken into consideration with

a covariance matrix and observations on different subjects are assumed to be independent. This is of importance because measures taken on the skull surface can be highly correlated or not, depending on the specific processes occurring in the "brain" under the electrode locations at the time of recording (Bullock et al., 2000; Thatcher, 2001). Since individual subjects are considered in a univariate analysis, group becomes another fixed effect along with frequency band and electrode location.

The general formulation of such a model is:

$$Y = X\beta + Z\gamma + \varepsilon$$

where X is a matrix whose columns represent the predictor variables associated with the fixed effects, β is the vector of fixed effect parameters, Z is a matrix whose columns represent the predictor variables associated with random effects, γ is the vector of random effects assumed to be randomly generated from a multivariate normal distribution, and β is the random vector of the errors. This equation represents the data vector for the mean where $E(\gamma) = 0$, $E(\varepsilon) = 0$, $V(\gamma) = G$ and $V(\varepsilon) = R$. Therefore $E(Y) = X\beta$ and by assuming that γ and ε are independent, the covariance of Y is modeled as:

$$V(Y) = ZGZ' + R$$

In our approach we disregarded potential random effects that would not be specific to single individuals. Thus, by absorbing potential within-subject random effects into the covariance matrix R, we can work with the simplified model

$$V(Y) = R$$

As a complementary approach we used a cross-validation algorithm based on the jackknife principle (i.e., it fits the model under consideration to the n data sets obtained by alternatively excluding each of the n subjects). For each of these n models, the sum of the squared prediction errors across sites and frequency bands at the omitted subject is computed. The mean of these sums across the n models (i.e., the mean multivariate prediction error) was considered as a measure of the misfit of the underlying model. According to the cross-validation 'philosophy,' the best model is the one with the lowest mean square error (MSE). A model containing too many variables or generally too many

degrees of freedom tends to over fit the underlying data, thereby picking up features that are specific to the data set from which it has been derived. Since these features can not be generalized, such a model will not perform very well on new data. The opposite is a model that is too coarse and does not describe the general features of the data well enough. In both cases the MSE obtained by cross-validation would be larger than one would want. On the other hand, the model minimizing the MSE generally has the advantage of representing a good compromise between these two extremes, in that it neither over nor under fits the underlying data. Therefore it captures the essential features of the data and tends to perform better in future predictions.

The amplitude and power models considered were of the form:

$$E(Y|site) = (1+group+group^2)*(1+site+site^2)_1$$

where group and site are ordinal variables with 5 and 4 levels, respectively.

The right side of the equation was given in symbolic notation (i.e., pairwise multiplication of terms within the two parentheses provides the predictor terms for the model), without their regression coefficients. In this model, containing eight fixed effect parameters and four observations per subject (one for each of the four sites), the R matrix is block-diagonal with blocks of dimension four.

For coherence data, a model of the following form was assumed for each frequency band:

$$E(Y|sitea,siteb) = (1+group+group^2)*(1+sitea+sitea^2+siteb+siteb^2+site*siteb)$$

where group, sitea and siteb represent ordinal variables with 5, 4 and 4 levels, respectively.

Again, symbolic notation is used to define the predictor terms of the model. In this model, containing 17 fixed effect parameters and 16 observations per subject (one for each pair of sites from the left and right hemisphere), the R matrix is block-diagonal with blocks of dimension 16. The coherence data of these 16 pairs were considered to be a function of two variables, the left hemispheric site (*sitea*) and the right hemispheric site (*siteb*). Intra- and interhemisphere coherence data were examined. A direct inclusion of the band variable into all models overcharged them; therefore, each band was evaluated separately.

An example using an absolute power data set will be used to show one way to apply the mixed model to qEEG data in clinical research.

The data sample is from a small set of five groups of individuals in a study looking at psychiatric disorders. The analysis used the mean from eight electrode locations (F3, F4, P3, P4, T5, T6, O1, O2) and 50 frequency bins (0 to 25 hertz in 0.5 intervals). These bins were converted to seven frequency bands (*band*); subdelta–0.5 to 1, delta–1 to 3.5, theta–4 to 7.5, alpha1–7.5 to 10, alpha2–10 to 12.5, beta1–12.5 to 15, and beta2–15 to 25. Two ordinal variables; *group* (with values 1 to 5), *site* (with values 1 to 4 representing frontal, parietal, temporal and occipital locations) were considered along with their squares and interaction terms. In this example the power data represents the symmetry number that results when the right electrode Fourier data is subtracted from the left (l − r). The activity was converted to a standard power representation using Scan™ 4.2 software (Neuroscan Labs). Results from a GLM (Table 1), a mixed procedure (Table 2), and the MSE from the cross-validation procedure (Table 3), for band 5 or the 10 to 12.5 Hz alpha are presented for comparison and discussion. At this $point_1$ conventional backwards selection would occur using the mean squared error (MSE) as an additional 'external' criterion for the goodness of fit. Appendix A details the SAS program used to run the GLM, the mixed procedure and the cross-validation process.

A more refined backwards selection strategy might additionally include a criterion limiting change in parameter estimates (i.e., a variable whose omission from the model would cause a change in one of the remaining parameters beyond a predefined tolerance band would then be kept in the model irrespective of its statistical significance).

The first step when evaluating this model for significant variation in group-specific patterns is to begin with the *grpsite* variable as this statistic measures differences between groups at different electrode locations. If this term is significant then the influence of the site variable depends on the group to which the subject belongs. This allows the characterizing of group-specific average patterns. The methodology of mixed models also offers the possibility of estimating individual random effects, thereby allowing identification of subjects who strongly deviate from their group average. In this example there appears to be a group difference within the individual population that is nonspecific to the site of the activity. There was a significant *grpsite* difference in both analysis methods (see Tables 1 and 2) but the significant chi-square statistic seen in Null Model Likelihood Ratio Test (NMLRT) with the mixed procedure indicates the appropriateness of modeling the covariance structure (The MIXED Procedure). If this statistic is not significant, the GLM results should be used.

TABLE 1. GLM~Power Symmetry~Band 5

Dependent Variable: y1

Source	DF	Squares	Sum of Mean Square	F Value	Pr > F
Model	8	0.36800463	0.04600058	1.98	0.0549
Error	115	2.67084299	0.02322472		
Corrected Total	123	3.03884762			

R-Square	Coeff Var	Root MSE	y1 Mean
0.121100	2539.823	0.152397	0.006000

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	−0.5796111604	0.32838349	−1.77	0.0802
Group	0.6260315352	0.29124667	2.15	**0.0337**
grp2	−0.1028814672	0.0486608	−2.11	0.0367
site	0.5773199036	0.29957828	1.93	0.0564
site2	−0.0975468	0.05897942	−1.65	0.1009
grpsite	−0.6155785328	0.26569904	−2.32	**0.0223**
grpsite2	0.1086477571	0.05230945	2.08	0.04
grp2site	0.0997438742	0.04439237	2.25	0.0266
grp2sit2	−0.0177132003	0.00873974	−2.03	0.045

TABLE 2. Mixed GLM~Power Symmetry~Band 5

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
9	30.26	**0.0004**

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	−0.4719	0.2493	28	−1.89	0.0687
Group	0.5077	0.2211	28	2.3	**0.0293**
grp2	−0.08247	0.03694	28	−2.23	0.0337
site	0.4954	0.2782	28	1.78	0.0858
site2	−0.08397	0.05922	28	−1.42	0.1672
grpsite	−0.5256	0.2467	28	−2.13	**0.042**
grpsite2	0.09374	0.05252	28	1.78	0.0851
grp2site	0.08422	0.04122	28	2.04	0.0505
grp2sit2	−0.01514	0.008776	28	−1.73	0.0955

TABLE 3. Cross-Validation~Power Symmetry~Band 5

Mean squared error in cross validation

Analysis Variable: e2

N	Mean	Std Dev	Minimum	Maximum
31	0.1025276	0.0761427	0.0136425	0.2813329

The relationship of each individual's activity is considered to be independent; therefore; the difference in the significance of the results, between the GLM and the mixed procedure, implies that there was enough covariance within each individual's data set to account for this change. In this example, the use of a mixed general linear analysis, with its modeling of the covariance structure from each subject's data, resulted in a more accurate understanding of the data. Though in this case it might have seemed to be just a manipulation in or refinement of analysis, which may or may not be relevant to understanding the data, in other instances the difference could be crucial. Tables 4, 5, and 6 show the results from the analysis of interhemispheric coherence relationships in band 3 (4.0-7.5 Hz). From the GLM there appears to be a highly significant relationship between the left and right hemispheres but there does not appear to be group involvement reducing the importance of this finding. The significant NMLRT indicates the appropriateness of the mixed procedure and in this case reveals significance in the data that were obscured in the GLM which missed the unique individual electrophysiological patterns analyzed by an unstructured covariance matrix.

TABLE 4. GLM~Interhemisphere Coherence~Band 3

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.8685211204	0.27490749	3.16	0.0017
Group	−0.0066450178	0.24381827	−0.03	0.9783
Sitea	−0.2011315623	0.16182495	−1.24	0.2145
Siteb	−0.2655712713	0.16182495	−1.64	0.1014
grp2	0.0014740373	0.04073658	0.04	0.9712
sitea2	−0.0051468363	0.02964404	−0.17	0.8622
siteb2	0.007155804	0.02964404	0.24	0.8094
sitasitb	0.0830425769	0.02371523	3.5	**0.0005**
grpsitea	−0.0227375174	0.14352421	−0.16	0.8742
grpsiteb	0.0267065301	0.14352421	0.19	0.8525
grpsita2	0.0126972631	0.0262916	0.48	0.6294
grpsitb2	0.0043052348	0.0262916	0.16	0.87
grsiasib	−0.0134738101	0.02103328	−0.64	**0.5221**
grp2sita	0.0031302375	0.02397968	0.13	0.8962
grp2sitb	-0.0048079142	0.02397968	−0.2	0.8412
grp2sia2	-0.0018778629	0.00439274	−0.43	0.6692
grp2sib2	-0.0006365416	0.00439274	−0.14	0.8848
g2siasib	0.0021117926	0.00351419	0.6	0.5482

TABLE 5. Mixed GLM~Interhemisphere Coherence~Band 3

Null Model Likelihood Ratio Test:

DF	Chi-Square	P > ChiSq
135	1035.32	**< 0.0001**

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	1.1125	0.1055	28	10.54	< 0.0001
Group	−0.01433	0.09359	28	−0.15	0.8794
Sitea	−0.1382	0.05031	28	−2.75	0.0104
Siteb	−0.3117	0.03927	28	−7.94	**< 0.0001**
grp2	0.002705	0.01564	28	0.17	0.8639
sitea2	−0.02204	0.007748	28	−2.84	0.0082
siteb2	0.009517	0.00633	28	1.50	0.1439
sitasitb	0.06914	0.007816	28	8.85	**< 0.0001**
grpsitea	0.03569	0.04462	28	0.80	0.4306
grpsiteb	0.05182	0.03483	28	1.49	0.148
grpsita2	−0.00113	0.006872	28	−0.16	0.8705
grpsitb2	−0.0036	0.005614	28	−0.64	0.5262
grsiasib	−0.01653	0.006932	28	−2.39	**0.0241**
grp2sita	−0.0056	0.007456	28	−0.75	0.4589
grp2sitb	−0.00883	0.00582	28	−1.52	0.1405
grp2sia2	0.00017	0.001148	28	0.15	0.8833
grp2sib2	0.000642	0.000938	28	0.68	0.4993
g2siasib	0.002689	0.001158	28	2.32	0.0278

TABLE 6. Cross-Validation~Interhemisphere Coherence~Band 3

Mean squared error in cross-validation

Analysis Variable: e2

N	Mean	Std Dev	Minimum	Maximum
496	0.0234607	0.0367621	1.5483183-9	0.2356986

What does all this mean? The more accurate the model, the more precise the prediction. Understanding the unique neurophysiological pieces that produce complex behaviors leads to more efficient intervention strategies. To help do this we create complex statistical models and then evaluate data sets for the best model that predicts the data. By increasing the complexity of statistical analysis possible, computers have allowed models to be created that represent many more of the basic physiological pieces that make up the total information available in a quantitative electroencephalogram. This modeling of complexity in

analysis becomes more and more analogous to modeling the brain's underlying neurophysiology processes. As data sets become larger, the ability to classify neurophysiological information will continue to improve. Clinically, being able to predict which individual will exhibit destructive negative behavioral patterns and act on them is still in the future, yet today we can predict some behavioral characteristics and are intervening clinically either before or when the predicted behavior becomes a problem. The use of a mixed procedure and its method of modeling the data structure appears to provide a more accurate and objective method of analysis which also provides quantifiable equations for testing predictions. Other methods, such as factor and cluster analysis, may need to be applied as we continue to increase our understanding of what is represented in the electroencephalogram.

It is hoped that the way the statistical model is presented in this paper will allow clinicians to be able to easily modify available SAS programs for use in understanding the data they gather in their daily practices. All of the SAS programs that are used in this paper are available by e-mail from the first author at *dbars2001@yahoo.com/*.

REFERENCES

Bullock, T. H., McClune, M. C., Achimowicz, J. Z., Iragui-Madoz, V. J., Duckrow, R. B., & Spencer, S. S. (2000). EEG coherence has structure in the millimeter domain: Subdural and hippocampal recordings from epileptic patients. *Electroencephalography and Clinical Neurophysiology, 95,* 161-177.

Cnaan, A., Laird, N. M., & Slasor, P. (1997). Tutorial in biostatistics: Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16,* 2349-2380.

Koch, G. G., Tangen, C. M., Jung, J. W., & Amara, I. A. (1998). Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine, 17,* 1863-1892.

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Tutorial in biostatistics: Modeling covariance structure in the analysis of repeated measures data. *Statistics in Medicine, 19,* 1793-1819.

SAS® system. SAS OnlineDoc™, Version 8, Chapter 41: The MIXED Procedure (MIXED PROC). *http://v8doc.sas.com/sashtml*.

Thatcher, R. (2001). An EEG severity index of traumatic brain injury. *Journal of Neuropsychiatry and Clinical Neuroscience, 13* (1), 77-87.

APPENDIX

```
FILENAME master Here you need where the data file is located on your computer, e.g.,
'c:\My Documents\My SAS Files\SASwork\fftpowermaster.txt';
  options nocenter ls=75;
%macro bla;
%let model =  group grp2 site site2 grpsite grpsite2 grp2site grp2sit2;
data master;
  infile master dlm='09'X dsd missover lrecl=1500;

input Fp1 Fp2 F3 F4 C3 C4 P3 P4 O1 O2 F7 F8 T3 T4 T5 T6 Fz Cz Pz ind group cyc &&;
The variables that are in your data set (.txt file) need to be listed here, in the order they are
found in the file.

Fp1  =  log(Fp1);
Fp2  =  log(Fp2);
F3   =  log(F3);
F4   =  log(F4);
C3   =  log(C3);
C4   =  log(C4);
P3   =  log(P3);
P4   =  log(P4);
O1   =  log(O1);
O2   =  log(O2);
F7   =  log(F7);
F8   =  log(F8);
T3   =  log(T3);
T4   =  log(T4);
T5   =  log(T5);
T6   =  log(T6);
Fz   =  log(Fz);
Cz   =  log(Cz);
Pz   =  log(Pz);

proc sort;
  by group ind cyc;
proc transpose out=o prefix=y;
     var Fp1 Fp2 F3 F4 C3 C4 P3 P4 O1 O2 F7 F8 T3 T4 T5 T6 Fz Cz Pz;
     by group ind cyc;

data o; set o;
  subdelta=(0.5<=cyc<=1.0);
  delta=(1.0<=cyc<=3.5);
  theta=(4<=cyc<=7.5);
  alpha1=(7.5<cyc<=10);
  alpha2=(10<cyc<=12.5);
```

```
  beta1=(12.5<cyc<=15);
  beta2=(15.5<cyc<=25);

band=subdelta+2*delta+3*theta+4*alpha1+5*alpha2+6*beta1+7*beta2;
  if _name_ in ('F3' 'F4') then site=1;
  else if _name_ in ('P3' 'P4') then site=2;
  else if _name_ in ('T5' 'T6') then site=3;
  else if _name_ in ('O1' 'O2' ) then site=4;

********** activated for comparison between hemispheres **********;
    if _name_ in ('F3' 'P3' 'T5' 'O1') then y1=-y1;
*******************************************************************************;

  if band=0 then delete;
proc summary nway;
  var y1;
  class group ind band site;
  output out=o2 mean=y1;

data o2; set o2;
  bandsite=band*site;
  grpband=group*band;
  grpsite=group*site;
  grpbasit=group*bandsite;
  band2=band**2;
  site2=site**2;
  grpband2=group*band2;
  grpsite2=group*site2;
  grp2=group**2;
  grp2site=grp2*site;
  grp2band=grp2*band;
  grp2ban2=grp2*band2;
  grp2sit2=grp2*site2;
  grp2basi=grp2*bandsite;

*******************************************************************************;
  /*   This will set the band for analysis */
*******************************************************************************;

  if band=5;

*******************************************************************************;
  /*   This will accomplish the glm procedure */
*******************************************************************************;
proc glm ;
 TITLE ' ###  GLM ### ';
```

```
  class ind;
  model y1 = &model / solution;

*************************************************************************;
  /*   This will accomplish the mixed procedure */
*************************************************************************;

proc mixed IC;
 TITLE ' ###  Mixed GLM ### ';
  class ind;
  model y1 = &model / solution;
  repeated / type=un subject=ind;

*************************************************************************;
  /*   This will accomplish the cross-validation procedure */
*************************************************************************;

data t; set o2;
  i=ind;
%do i=1% to 31;
  data t; set t;
     z=y1;
     if i=&i then z=.;
  %if &i ne 1 %then %do;
     data t; set t;
      drop pred;
  %end;
  proc reg data=t noprint;
     TITLE ' ### Cross-validation ### ';

  model z = &model;
       output out=t p=pred;
  data t; set t;
     if i=&i then e2=(y1-pred)**2;
%end;

%mend bla;
%bla;

proc summary data=t nway;
     var e2;
     class ind;
     output out=s sum=s_e2;
proc means data=s;
     var s_e2;
     title 'Mean squared error in cross-validation';
run;
```