Model Selection in Linear Mixed Effects Models Using SAS® PROC MIXED

Long Ngo, Ischemia Research, San Francisco ; University of California, Berkeley, CA
Richard Brand, University of California, Berkeley, CA

Abstract

Although there are disadvantages associated with model building procedures such as backward, forward and stepwise procedures (e.g. multiple testing, arbitrary significance level used in dropping or acquiring variables), many analysts use these procedures and are not aware that alternative modeling selection methods exist. This paper focuses on model selection using the Akaike Information Criteria (AIC) in the case of linear mixed-effects models. AIC's fundamental concepts are reviewed and two examples are given to demonstrate its use through PROC MIXED. Master-level biostatisticians, epidemiologists, and others who are working with longitudinal data are encouraged to investigate AIC as the tool in modeling repeated measures data.

1.0 Introduction

Model selection is one of the most frequently encountered problems in data analysis. In most observational epidemiological studies, investigators frequently attempt to construct the most desirable statistical model using the popular methods of forward, backward, and stepwise regression (4). Of course knowledge of the subject matter plays an important role in model selection, but if based strictly on the data, model selection is often carried out using one of the automated procedures built into the software, of which the most popular method is perhaps stepwise model selection. These methods pose the problem of the arbitrary selection of the significance level(s) in allowing a variable to enter into or to be dropped from the model during the selection process (1,4). There is also the problem of multiple testing that comes with fitting and refitting the model (1,4). The issue is made more complicated in the case of repeated or longitudinal data where selecting the best model means not only to select the best mean structure but also the most optimal variance-covariance structure (10,11). This paper reviews another model selection method which helps eliminate the problems associated with setting an arbitrary significance level required in automated procedures such as stepwise. Using a criterion like AIC for selecting a model, bypasses the need to specify a significance level in a model building process. The Akaike Information Criteria (AIC) and its principles in model selection will be described. The paper will also show how one can set up a model selection strategy using AIC in linear mixed-effects model framework. AIC and other related criteria (SBC, HQ, CAIC) are fully available in SAS PROC MIXED.

Real data from two case studies will be presented to show how model selection using AIC was used to achieve the desired objective.

2.0 Some Fundamental Concepts of AIC

In order to understand the principle behind AIC, one needs to return to the definition of the Kullback-Leibler information (5,8) which is considered to be a measure of the distance between two density functions. In a model selection problem, one would like to select the model family which performs best. The distance between the true model and the selected model can be represented by the Kullback-Leibler information. If one assumes the true model's density to be $f(.)$ and the joint density function for the selected model to be $g(., \hat{\theta})$, where $\hat{\theta}$ is the estimated vector of d parameters by the maximum likelihood method, then the Kullback-Liebler distance can be written as

$$\Delta_{K-L}(f(.), g(., \hat{\theta})) = \int \log \frac{f(x)}{g(x, \hat{\theta})} f(x) dx \quad (*)$$

where x is the observed sample data of n independent observations. In this setup, the expectation of (*) provides the basis for model selection, and the estimate of this expectation provides a criterion for model selection. The asymptotic approximation for the estimation of this expectation is given as (6,8)

$$\hat{E}\left(\Delta_{K-L}(f(.), g(., \hat{\theta}))\right) = \int f(x) \log(f(x)) dx + (-\log(g(x, \hat{\theta}))) + trace(\hat{\Sigma} \hat{\Omega}^{-1}) \quad (1)$$

where,

$\log g(x, \hat{\theta})$ is the maximized log likelihood function based on the observed data, and $\hat{\theta}$ is the maximum likelihood estimate of the parameter vector θ, and

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{d \log g(x_i, \hat{\theta})}{d\theta_u} \right]\left[\frac{d \log g(x_i, \hat{\theta})}{d\theta_v} \right], u, v = 1, 2, ..., d,$$

and

$$\hat{\Omega} = \frac{-1}{n} \sum_{i=1}^{n} \left[\frac{d^2 \log g(x_i, \hat{\theta})}{d\theta_u d\theta_v} \right], u, v = 1, 2, ..., d$$

Notice that the term $\int f(x)\log(f(x))dx$ is dependent on the true model which is unknown; however, this term is fixed when comparing between models. The part of this estimated expectation of the Kullback-Liebler distance that is needed in the comparison process is

$$-\log g(x,\hat{\theta}) + trace(\hat{\Sigma}\,\hat{\Omega}^{-1}) \quad (2)$$

which is computable for the parametric model.

When one can assume that the true model is contained within the family of models from which the fitted model is obtained then one can write (2) as $-\log g(x,\hat{\theta}) + d$ (3). AIC is basically twice the expression of (3) so AIC = -2loglikelihood + 2d (4). Thus AIC imposes a penalty of two units per parameter in the model.

Based on AIC , as the model selection criterion, then among all possible models considered, the one with the smallest value of AIC is considered to be the best model. Notice that if the true model $f(.)$ is very different (not contained in the family model that generates the selected model) from $g(.,\theta)$, then one should compute the trace term of $\hat{\Sigma}\,\hat{\Omega}^{-1}$ directly which is more difficult and computationally expensive.

3.0 Model selection using AIC for linear mixed-effects models

Suppose there are n independent subjects with m_i correlated measurements, $i = 1,2,\ldots,n$.
The marginal probability density of y_i is

$$f(y_i|b,\Sigma_i(\phi)) = \frac{1}{(2\pi)^{\frac{m_i}{2}}|\Sigma_i(\phi)|^{-\frac{1}{2}}}\exp(-\frac{1}{2}(y_i - x_i b)'\Sigma_i(\phi)^{-1}(y_i - x_i b))$$

The vector of both fixed and random parameters is $\theta = (\phi,b)$. So the maximized log likelihood is then

$$\log f_{\hat{\theta}}(y_i) = -\frac{m_i}{2}\log 2\pi + \log|\Sigma_i(\hat{\phi})|^{-\frac{1}{2}} - \frac{1}{2}(y_i - x_i\hat{b})'|\Sigma_i(\hat{\phi})|^{-1}(y_i - x_i\hat{b})$$

From here AIC can be computed easily with d being the sum of both the fixed and random effects parameters. The SAS system version 6.11 gives this value AIC for both the maximum likelihood and the restricted maximum likelihood case (2). One can also compute the trace term

of $\hat{\Sigma}\,\hat{\Omega}^{-1}$ given in (2) since the density function of the fitted model is known.

There are two different methods presented here for selecting the "best" or the "smallest AIC" model among all the models under consideration. The first method essentially identifies all the possible mean functions and all the possible variance-covariance structures applicable to the question of interest. The number of possible models for consideration includes all combinations resulted from both the mean and variance-covariance structures. For each of these models, AIC is then computed and the model with minimum AIC is selected.

The second method recommended by Wolfinger and Diggle (1,2,10) has the following steps : First, using the most complex mean structure under consideration, select the best variance-covariance structure using the restricted maximum likelihood (REML). REML which focuses on the covariance side of the model should be used in place of the likelihood. Let AIC_R denotes AIC derived from REML : AIC_R = -2*(Restricted likelihood) + 2*(# covariance parameters). The variance-covariance structure with the smallest AIC_R is selected. Then using this variance-covariance structure, go back and use AIC to select the best mean structure. In the following example, both methods will be illustrated.

4.0 Example 1

The data are from 1006 subjects who participated in an epidemiology study on aging. Each subject was measured every minute for 3 to 30 minutes (TIME) from the start of a treadmill test. The response called VO2KG, which measures physical fitness, is the amount of expended energy from exercise. The higher the VO2KG, the better the physical performance. DURATION stands for the total time (minutes) the subject was able to exercise. AGE is in years. The analysis is done for each gender. This is a repeated measures study where the objective is to construct the normogram shown in figure 1. Table 1 shows all the mean structures and variance-covariances of interest. A total of 18 models were fitted, each has a AIC value. For each of 18 models considered, the rounded AIC values are given in table 1.

Table 1 : AIC values for all models considered

	TIME	DURA TION	AGE	TIME+ DURA TION	TIME + AGE	**TIME + DURA TION+ AGE**
Simple	80201	96706	97366	78741	79899	78683
CS	65320	94723	94768	65041	65118	65034
AR(1)	61706	63680	63700	61392	61525	**61384**

The minimum value of AIC=61384 from the model with mean structure of TIME+DURATION+AGE and variance-covariance AR(1) indicates the best model among the 18 considered. The predicted value of the response VO2KG was computed from this model and then smoothed to allow the construction of the normogram which the physician can use to classify the patient into the appropriate percentile category. Figure 1 showed the normogram for the male group constructed by using the predicted VO2KG from the selected model.

Figure 1 : Smoothed Percentile Chart for Male

5th, 25th, 50th, 75th, 95th Percentile
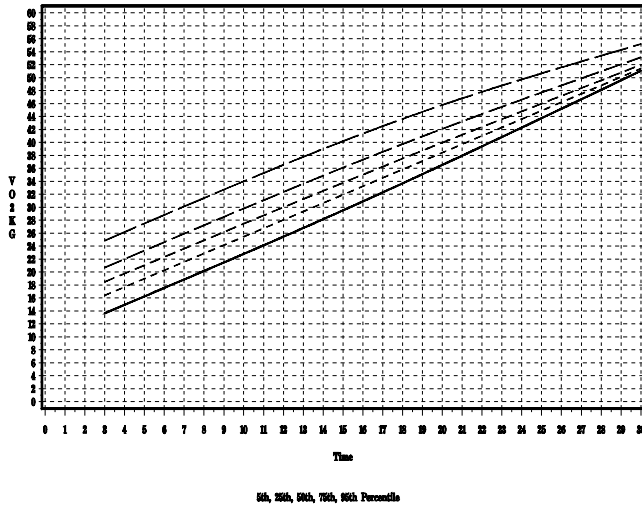
Using method 2, the model with the most complex mean structure among those considered is the one with TIME+DURATION+AGE. This model is fitted using the three variance-covariance structures and the AIC_R is minimum for AR(1). Then using AIC to reduce the mean structure. In this case AIC chooses the mean structure TIME+DURATION+AGE so the model selected is exactly the same as the one chosen in method 1.

5.0 Example 2

Data collected on 9 individuals, each had 3 visits (one or two weeks apart, except for one subject with one visit, and one other subject with two visits). For each visit, one measurement was made on each eye to assess the permeability (Pdc) to sodium fluorescein of the corneal epithelium. The objective is to obtain the variance component estimates from the best model. There are five mean structures of interest (intercept, eye, visit, eye+visit, eye+visit+eye*visit) and five variance-covariance structures of interest (simple, compound symmetry, compound symmetry+eye, compound symmetry+visit, compound symmetry+eye+visit) resulting in 25 models. In this description of the candidate models, compound

symmetry+eye for example, means that the R matrix in PROC Mixed specification is of type compound symmetry and 'eye' is the random effect. Figure 2 shows the AIC for all 25 models, and the lowest AIC (the best model), is from the model with 'eye' as the only fixed effect, and compound symmetry as the variance-covariance structure.
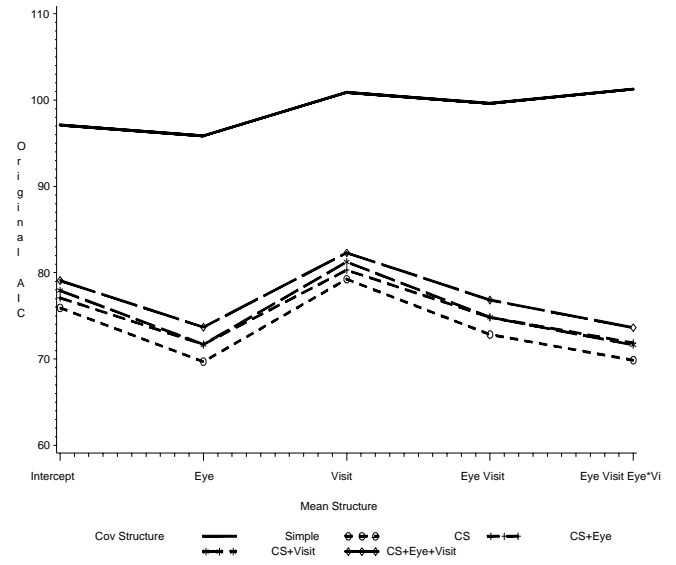
Appendix 1 shows the SAS macro program for automating the selection procedure for this example. Also using the second method, the same model was chosen as the best model.

6.0 Additional issues

The AIC is known to be inconsistent (1,8,9). Other criteria such CAIC (1), SBC (7), HQ (3) are consistent criteria which are also available in SAS 6.11. Model selection using formula (2) which has the trace term may have better consistency property than the AIC but this has not been implemented. Using simulation, we are currently evaluating the performance and consistency of the criterion using the trace term formula (2). Linhart and Zuchini (7) also discuss other discrepancy measures which emphasize on the fixed effects side of the linear mixed-effects model. In certain practical situations, these may be preferable to the Kullback-Liebler based criteria.

7.0 Conclusion

This papers reviews some fundamental concepts of the AIC and shows how model selection for linear mixed-effects models can be done using AIC. Analysts are encouraged to utilize selection criteria such as AIC and others which are already implemented in the SAS software. We presented

two different procedures for selecting the best model. The procedure suggested by Diggle and Wolfinger of using REML for variance-covariance structure selection, and ML for mean structure selection has a computational advantage. The number of possible models evaluated for this procedure is equal to the sum of the mean structures and variance-covariance structures in consideration. The method we proposed using all combinations of mean structures and variance-covariance structures of interest is more computationally expensive especially when the number of mean structures and/or variance-covariance structures is large. However, this method provides a more comprehensive examination of all models evaluated.

8.0 Reference

1. Bozdogan, H. (1987). Model selection and Akaike's information criteria (AIC) : the general theory and its analytical extensions. Psychometrika 52, 345-370.
2. Diggle, P; Liang, K; Zeger, S. (1994). Analysis of Longitudinal Data, Oxford Press.
3. Hannan, E.J. and Quinn, A.G. (1979). The determination of the order of an autoregression. Journal of Royal Statistical Society, Vol. B 41, 190-195.
4. Hosmer, D.W. and Lemeshow, S. (1989). Applied Logistic Regression. New York : John Wiley and Sons, Inc.
5. Kullback, Solomon (1978). Information Theory and Statistics, Massachusetts: Peter Smith.
6. Linhart, H and Zuchini, W. (1986). Model Selection. New York : John Wiley and Sons, Inc.
7. Schwarz, G. (1978). Estimating the dimension of a model. Annual of Statistics, Vol. 6, 461-464.
8. Shibata, Ritei (1989). From Data to Model. New York : Springer-Verlag, 215-240.
9. Shibata, Ritei (1986). Consistency of model selection and parameter estimation. Essays in Time Series and Allied Processes. Applied Probability Trust, 127-141.
10. Wolfinger, R.D. (1993). Covariance structure selection in general mixed models. Communications in Statistics, Simulation and Computation, Vol. 22, 1079-1106.
11. Wolfinger, R.D. (1996). Heterogeneous variance-covariance structures for repeated measures. Journal of Agricultural, Biological, and Environmental Statistics, Vol. 1, Number 2, 205-230.

9.0 Acknowledgements

10.0 Authors

Long Ngo
University of California, Berkeley
140 Warren Hall
Berkeley CA 94720
(510) 643-5716
lhn@orion.iref.org

Richard Brand, Ph.D.
University of California, Berkeley
Haviland Hall
Berkeley CA 94720
(510) 642-4627
brand@stat.berkeley.edu

Appendix 1

```
/*********************************************
Program  :    ex5.sas
Time     :    5-16-96 3:33:07 pm
By       :    Long Ngo
Input    :    e:\brand\pcerept\saslib\pcerept2.sd2
Output   :
Purpose  :           Automating the macro
                     for model selection for example 2
                     implementing adjustment for AIC and
                     consistent AIC estimate.
*********************************************/
options ls=80 ps=60 pageno=1 mprint;;
libname rjb 'e:\brand\pcerept\saslib';
title1 'ex5.sas';
data a1;
  set rjb.pcerept2;
  format _all_;     *strip all defined formats off the data;
run;
proc sort;
  by id;
run;
data rjb.aicinfo;    *establish the final data set structure;
  length mean $25 randvar repttype $15 criteria $3 aic
nobs nfixed ncov nsubj 5;
run;
%macro v(mean,repttype);  * macro for handling just
fixed effects;
proc mixed data=a1 method=ml;
class id eye visit;
model lnpce = &mean ;
repeated / type=&repttype subject=id;
title2 "Model with Mean Structure = &mean ";
title3 "Repeated Type = &repttype";
make 'fitting'out=aout;
make 'SolutionF' out=aoutfix;
make 'CovParms' out=aoutcov;
run;
%f;
%mend;
%macro vr(mean,randvar,repttype);  * macro for handling
both fixed and random effects;
proc mixed data=a1 method=ml;
class id eye visit;
model lnpce = &mean ;
random &randvar / subject=id;
repeated / type=&repttype subject=id;
title2 "Model with Mean Structure = &mean ";
title3 "Random Effects = &randvar";
title4 "Repeated Type = &repttype";
make 'fitting'out=aout;
make 'SolutionF' out=aoutfix;
make 'CovParms' out=aoutcov;
run;
%f;
%mend;

%macro f;
data aicinfo;
  length mean $25 randvar repttype $15;
  set aout;
  mean="&mean";
  randvar="&randvar";
  repttype="&repttype";
  where descr=:'Akaike';
  aic=value;
run;
data nobs;
  set aout;
  where descr=:'Observation';
  nobs=value;  *obtain number of observations in analysis;
run;
data tfix1;
  set aoutfix (keep=est);
  if est ne 0;
run;
proc univariate noprint;
  var est;
  output out=nfixed n=nfixed;  *get the number of fixed
effect parameters;
run;
data tcov;
  set aoutcov (keep=est);
run;
proc univariate noprint;
  var est;
  output out=ncov n=ncov;  *get the number of covariance
parameters;
run;
*get the number of available subjects for each model;
data nsubj;
  set rjb.pcerept2 (keep=id);
run;
proc sort;
  by id;
run;
data nsubj;
  set nsubj;
  by id;
  if first.id;
  c=1;
run;
proc univariate noprint;
  var c;
  output out=nsubj sum=nsubj;
run;
* construct the desired data set;
data aicinfo;
  merge aicinfo nfixed nobs ncov nsubj;
run;
data rjb.aicinfo;
  set rjb.aicinfo aicinfo;
run;
%mend;
```

5

```
*(mean repeated type for macro %v) and (mean random
repeated type for %vr);

%v(/ s,simple);
%v(/ s,cs);
%vr(/ s,eye,cs);
%vr(/ s,visit,cs);
%vr(/ s,eye visit, cs);

%v(eye / s,simple);
%v(eye / s,cs);
%vr(eye / s,eye,cs);
%vr(eye / s,visit,cs);
%vr(eye / s,eye visit, cs);

%v(visit / s,simple);
%v(visit / s,cs);
%vr(visit / s,eye,cs);
%vr(visit / s,visit,cs);
%vr(visit / s,eye visit, cs);

%v(eye visit / s,simple);
%v(eye visit / s,cs);
%vr(eye visit / s,eye,cs);
%vr(eye visit / s,visit,cs);
%vr(eye visit / s ,eye visit, cs);

%v(eye visit eye*visit / s,simple);
%v(eye visit eye*visit / s,cs);
%vr(eye visit eye*visit / s,eye,cs);
%vr(eye visit eye*visit / s,visit,cs);
%vr(eye visit eye*visit / s,eye visit, cs);

proc format;
  value
        meantype
        1='Intercept'
        2='Eye'
        3='Visit'
        4='Eye Visit'
        5='Eye VisitEye*Visit';

  value covtype
        1='Simple'
        2='CS'
        3='CS+EYE'
        4='CS+Visit'
        5='CS+Eye+Visit';
run;
* set up the data for graphing;

data rjb.t1;
  set rjb.aicinfo;
  if randvar =: '&' thenrandvar='None';
  if _n_=1 then delete; *drop the null case used for setting;
  if mean='/ s' then do;
    mean='Intercept';
    meantype=1;
```

```
  end; else
  if mean='eye / s' then meantype=2; else
  if mean='visit / s' then meantype=3; else
  if mean='eye visit / s' then meantype=4; else
  if mean='eye visiteye*visit / s' then meantype=5;

  if randvar='None' andrepttype='simple' then covtype=1;
else
  if randvar='None' andrepttype='cs'    then covtype=2;
else
  if randvar='eye' andrepttype='cs'     then covtype=3; else
  if randvar='visit' andrepttype='cs'   then covtype=4; else
  if randvar='eye visit' andrepttype='cs' then covtype=5;
  if meantype=1 then ncov=ncov-1; *adjust for the resid;
  logl = aic+ncov ; *Mixed defines AIC = logl - number of
cov parameters;
                    *this is to get back logl value;
  aic = -2*logl + 2*(nfixed+ncov); *recompute original
def. of AIC;
  aic_con = -2*logl + (nfixed+ncov)*(log(nsubj)+1);
*consistent AIC estimate;
  label meantype='Mean Structure'
        covtype ='Cov Structure'
        aic = 'Original AIC'
        aic_con = 'Consistent AIC'
        logl = 'Log Likelihood'
        nobs = 'Number ofObs in Analysis'
        nsubj = 'Number of Subjects in Analysis'
        nfixed = 'Number of Fixed Effects Parameters'
        ncov = 'Number of COV parameters';
  format meantype meantype. covtype covtype.;
  drop descr value mean;
run;
proc sort ;
  by meantype covtype;
run;
goptions device = winprtm;
symbol1 c=black v=point i=j l=1 width=3;
symbol2 c=black v=circle i=j l=2 width=3;
symbol3 c=black v=plus i=j l=3 width=3;
symbol4 c=black v=star i=j l=4 width=3;
symbol5 c=black v=diamond i=j l=5 width=3;
axis1 order = (1 to 5) ;
proc gplot;
  plot aic * meantype = covtype / haxis=axis1;
title1 h=0.75 'Figure 1 : AIC Values Computed for 25
Models in Example 2';
run;
```