Paper PO20

A Parameterized SAS Macro to Select an Appropriate Covariance Structure in Repeated Measures Data Analysis Using PROC MIXED

Leela Aertker, Rho, Chapel Hill, NC
Paul Nguyen, Rho, Chapel Hill, NC
Charity Quick, Rho, Chapel Hill, NC

ABSTRACT

In clinical trials we often encounter multiple measurements on a subject or experimental unit over a period of time. When analyzing longitudinal data using SAS PROC MIXED, it is critical to specify the covariance or correlation structure of the repeated measures. Unless stated a priori, one must determine the most appropriate covariance structure by comparing and contrasting the best fit of different covariance structures based on AIC, AICc, BIC, or -2 log likelihood. This paper will describe an efficient macro to determine the covariance structure for a model. The %COVSTRUC macro includes several user-specified parameters: goodness-of-fit statistic, model information, and whether a particular covariance structure has priority given convergence. The covariance structure with the best fit (i.e., lowest value of specified fit statistic) will be chosen and outputted to a macro variable, which can then be called into the final analysis model.

INTRODUCTION

Longitudinal data or repeated measurements (e.g., vital signs, biomarkers) are frequently analyzed in clinical trials. When modeling repeated measures, one must determine the most appropriate covariance structure to obtain the best fit. If the covariance structure is not specified a priori, then one typically fits the model with several different covariance structures and manually compares the goodness-of-fit statistics (i.e., AIC, AICc, BIC, or -2 log likelihood).

SAS PROC MIXED provides the necessary criteria to determine the most appropriate covariance structure. The macro described within this paper uses the functionality provided within PROC MIXED and automates the comparison and selection of the covariance structure. The macro cycles though each covariance structure, outputs the specified goodness-of-fit statistics (i.e., AIC, AICc, BIC, or -2 log likelihood), and then determines which specified fit statistic is closest to 0. Lastly, a macro variable is outputted that specifies the covariance structure associated with the best fit to be used in the final analysis model.

If the covariance structure is pre-specified in the protocol and/or analysis plan, then the macro determines whether the pre-specified covariance structure converges. If the covariance structure does not converge, then the macro will not output a covariance structure to be used in the final analysis model. One may reference the log for explanatory notes (e.g., UN does not converge) and the output file for a comparison table contrasting the different covariance structures.

AVAILABLE DATA

Data from the Inner-City Anti-IgE Therapy for Asthma (ICATA) were downloaded from the Immunology Database and Analysis Portal (ImmPort) system, which was developed by the Bioinformatics Integration Support Contract (BISC) Phase II of the Northrop Grumman Information Technology Health Solutions team for the National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIAID), and the Division of Allergy, Immunology, and Transplantation (DAIT).

The ICATA study was a multi-center, double-blind, randomized controlled trial that compared the effect of omalizumab therapy and guidelines-based treatment with guidelines-based treatment alone in participants with moderate to severe asthma aged 6 to 20 years old. Participants visited the site every three months to complete questionnaires and biological measurements, including fractional exhaled nitric oxide (FeNO), an asthma inflammatory biomarker.

The dataset is structured as one record per participant per visit. There are multiple FeNO measurements for each participant within the 2 treatment groups: omalizumab therapy and guidelines-based treatment vs. guidelines-based treatment alone.

METHODS

With longitudinal data, we expect the residuals (i.e., unexplained portion of each subject's outcome) to be heteroscedastic and autocorrelated over time. If we adopt the homogeneity assumption, the entire error structure is repeated identically across subjects and allows for our distributional assumptions to be more parsimonious.

Kincaid (2005) and Singer & Willet (2003) recommend the following covariance structures when modeling longitudinal data in PROC MIXED:

- VC: Variance component is the default. It models a different variance component for each random or repeated effect.
- AR(1): First order autoregressive is a special case of TOEP where the variances are assumed to be homogeneous and the covariance or correlation decline exponentially with time. That is, variability in a measurement is constant regardless of when it is measured and 2 visit measurements taken closer together are more correlated than those taken farther apart.
- ARH(1): Heterogeneous counterpart to AR(1).
- TOEP: Toeplitz assumes that visit measurements taken next to each other have the same correlation, but do not necessarily have the same pattern as in AR(1).
- TOEPH: Heterogenous counterpart to TOEP.
- CS: Compound symmetry assumes homogeneous variances and constant covariance or correlation between measurements across time.
- CSH: Heterogeneous counterpart to CS.
- UN: Unstructured does not assume homogenous variances. It requires fitting the most parameters and, therefore, requires the most observations. It is also less generalizable given the complexity of the model.

MACRO

```
PROC MIXED data=&data;
        class &class;
        model &model;
        repeated &repeated / TYPE = &type sub= &sub r;
run;
%macro covstruc(data=, class=, model=, repeated=, type=, sub=,modelfit=);
```

DATA	Name of input dataset. This is a REQUIRED parameter and there is no default. User should restructure data as appropriate.
CLASS	Class statement for PROC MIXED. This is a REQUIRED parameter and there is no default. E.g., ID VISIT TRT.
MODEL	Model statement to be used in PROC MIXED. This is a REQUIRED parameter and there is no default. E.g., FeNO = VISIT TRT VISIT*TRT.
REPEATED	Repeated statement to be used in PROC MIXED. This is a REQUIRED parameter and there is no default. E.g., VISIT.
TYPE	Defines the pre-specified covariance structure. User can specify one type or not specify a type at all. The default is to compare all aforementioned covariance structures.
SUB	Subject identifier to be used in PROC MIXED. This is a REQUIRED parameter and there is no default. E.g., ID.
MODELFIT	Defines which fit criterion will be used. Must be expressed as the values -2LL, BIC, AICC, or AIC. This is a REQUIRED parameter and the default is AIC.

RESULTS

The example data includes a dependent variable (FeNO) and two independent variables (TRT and VISIT). A partial segment of our example dataset is as follows:

ID	TRT	VISIT	FeNO
101	1	1	10.8
101	1	2	10.6
101	1	3	9.8
102	2	1	20.9
102	2	2	20.0
102	2	3	20.3

Model Fitting Results:

Fit-Statistics

Fitting Criteria	Best Fit	Minimum Value	TOEP	UN	AR(1)
AIC (smaller is better)	UN	2426.9	2843.3	2426.9	2900.9

CONCLUSIONS

This validated macro is yet another tool for both programmers and statisticians alike, which efficiently selects the covariance structure associated with the best fit for easy inclusion in the final analysis model. It eliminates the tedium of comparing variance covariance structures and automates the selection process. The macro includes several covariance structures to consider in the repeated measures paradigm and employs goodness-of-fit statistics to choose the most appropriate covariance structures. However, goodness-of-fit statistics may result in conflicting results and are not always guaranteed to lead to the correct structure (Kincaid, 2005). That being said, there is no substitution for understanding your data.

REFERENCES

Kincaid, C. (2005). Guidelines for selecting the covariance structure in mixed model analysis. *Statistics and Data Analysis* 30:1-8.

Singer JD, Willet JB. (2003). *Applied Longitudinal Data Analysis*. New York, New York: Oxford University Press.

CONTACT INFORMATION

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

Leela Aertker
Senior Biostatistician, Rho
6330 Quadrangle Drive
Chapel Hill, NC 27514
(919) 595-6286
Leela_Aertker@rhoworld.com