The Journal of Animal & Plant Sciences, 22(3): 2012, Page: 552-555
ISSN: 1018-7081

DETERMINATION OF APPROPRIATE COVARIANCE STRUCTURES IN RANDOM SLOPE AND INTERCEPT MODEL APPLIED IN REPEATED MEASURES

Gazel Ser

Department of Animal Science, Faculty of Agriculture, Yuzuncu Yil University, 65100Van/ Turkey
Corresponding author e-mail: gazelser@yyu.edu.tr

ABSTRACT

This study aims to determine variance-covariance structures of dependent variable in data set containing repeated measures and to compare covariance parameter estimation methods. To this end, random intercept and slope model which is among the special cases of linear mixed model was formed and the time variable was involved into the model in a continuous and categorical manner. Also, compound symmetry (CS), toeplitz (TOEP), first-order autoregressive (AR(1)), homogeneous variance-covariance models and unstructured (UN), heterogeneous compound symmetry (CSH), heterogeneous toeplitz (TOEPH), heterogeneous first-order autoregressive (ARH(1)), first-order ante-dependence (ANTE(1)) and unstructured correlation (UNR) heterogeneous variance-covariance models were performed in order to determine the variance-covariance structure between the repeated measures. In addition, comparison of ML and REML was carried out as covariance parameter estimation method. Consequently, random intercept and slope model (RISM) was found to be the most appropriate one in modeling the repeated measure data when ML was used as the parameter estimation method and UN, CSH, ARH(1), TOEPH, ANTE(1), UNR as the covariance models.

Key words: Covariance structures, repeated data, linear mixed model.

INTRODUCTION

In the test designs including repeated measures, it is possible to get different features (live-weight, height at withers, body length etc in the field of stockbreeding) from variable structure test units with repeated measures made in different times for same features (Tabachnick and Fidel 2001). It is essential to properly identify the variance-covariance structure among the data in the analysis of the repeated measures (Akbaş *et al.,* 2001, Orhan *et al.,* 2010; Eyduran and Akbaş 2010). At this point, structure of general linear model (GLM) which is based on that the repeated measures in the data structure are different from each other and have a homogeneous variance is quite impractical for modeling the repeated measures. Therefore, the linear mixed model which enables flexible modeling of variance-covariance matrix structure is recommended for the statistical analysis of the repeated measure data, without imposing any restrictive assumption on the correlated data structure of the repeated measures taken on the same experimental unit (Iyit, 2008).

In present study, random intercept and slope model was formed on the data having a repeated measure structure by involving continuous and categorical effect of the time factor into the model variable. A variable such as "time 2" was formed in order to involve the time into the model categorically (Doganay, 2007). The chest girth feature treated as the repeated measure was involved into the model as the dependent variable. To that end, compound symmetry (CS), toeplitz (TOEP), first-order autoregressive (AR(1)), homogeneous variance-covariance models and unstructured (UN), heterogeneous compound symmetry (CSH), heterogeneous toeplitz (TOEPH), heterogeneous first-order autoregressive (ARH(1)), first-order ante-dependence (ANTE(1)) and unstructured correlation (UNR) heterogeneous variance-covariance models were performed in order to determine the variance-covariance matrix structure of the dependent variable. Comparison of ML and REML was also carried out as covariance parameter estimation method. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) cohesion criteria were used to determine the appropriate variance-covariance structure using covariance parameter estimation methods.

MATERIALS AND METHODS

Chest girth measured every 2 weeks (14-day) from 57 Il de france lambs up to 6-months of age, as the animal material, was treated as dependent variable whereas birth type, dam age, sex and live weight of animals were involved into the model as independent variables.

General Linear Mixed Models: General linear mixed model is defined as follows:

$$Y = X\beta + Zu + e \quad (1)$$

β : $p \times 1$ vector of fixed effects. It takes its place as the fixed effect in the model

u : $q \times 1$ vector of random effects. It takes its place as the independent $(u \sim N(0,G))$

X and Z : respectively $n \times p$ and $n \times q$ are design matrixes for the fixed and random effects

e : random error term (Kincaid 2005).

The maximum likelihood (ML) and restricted maximum likelihood (REML) methods: Let α, V_i denote unknown parameters in covariance matrix. Accordingly, conditions on α indicate a closed form for ML estimator of β. That is,

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1} \sum_{i=1}^{N} X_i' V_i^{-1} Y_i \quad (2)$$

($u_i \mid Y_i$) the posterior distributions average of the random effects is used for the estimation of random effects in the data provided.

$$\hat{u}_i = D Z_i' V_i^{-1} \left(Y_i - X_i \hat{\beta} \right) \quad (3)$$

Either ML or REML is used to estimate α. ML (L_1) and REML (L_2) estimations are acquired using following equations.

$$L_1(\alpha; y_1, \ldots, y_N) = c_1 - \frac{1}{2} \sum_{i=1}^{N} \log|V_i| - \frac{1}{2} \sum_{i=1}^{N} r_i' V_i^{-1} r_i \quad (4)$$

$$L_2(\alpha; y_1, \ldots, y_N) = c_2 - \frac{1}{2} \sum_{i=1}^{N} \log|V_i| - \frac{1}{2} \sum_{i=1}^{N} \log\left| X_i' V_i^{-1} X_i \right| - \frac{1}{2} \sum_{i=1}^{N} r_i' V_i^{-1} r_i \quad (5)$$

$$r_i = y_i - X_i \left(\sum_{i=1}^{N} X_i' V_i X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' V_i^{-1} y_i \right)$$

and c_1, c_2 are appropriate invariants. Equations (4) and (5) are maximized using iterative numerical techniques such as Fisher Scoring and Newton-Raphson. $\hat{\alpha}_{ML}$ or $\hat{\alpha}_{REML}$ estimations takes place of the unknown α in the equations (2) and (3) (Antonio ve Beirlant 2007).

RESULTS AND DISCUSSION

Cohesion criteria results of variance-covariance structures which were obtained from RISM model are given in Table 1 and parameter estimates regarding fixed effects in Table 2. ML and REML estimate values of covariance parameters regarding the random effects in the structure of RISM are given in Table 3 for UN, CSH, ARH(1), TOEPH, ANTE(1) and UNR.

UN, CSH, ARH(1), TOEPH, ANTE(1) and UNR structures were determined as the best covariance models respectively from Table 1, in modeling the change between the repeated measure data regarding the chest girth for the ML and REML covariance parameter estimation methods. CS, AR(1) and TOEP were found to be the worst covariance structures in modeling.

Moreover, goodness of fit results indicated ML as the best covariance parameter estimation methods . Both AIC and BIC cohesion criteria showed a tendency to rank the heterogeneous variance-covariance models as the best model group for the repeated measure data regarding the chest girth.

As shown in Table 2, sex, dam age ($p<0.05$), birth type, time 2 and live weight variable ($p<0.0001$) were found significant in CS, AR(1) and TOEP structures, however, sex and dam age were found to be insignificant whereas birth type ($p<0.05$), time 2 and live weight ($p<0.0001$) were found to be significant in the CSH, ARH(1), TOEPH, UN, ANTE(1), UNR structures. Also, homogeneous F values were found to be higher rather than heterogeneous structures.

Having used Table 3, the percents (intra-class correlation) of the experimental effect included as random effect terms into the model to explain change in the structure of dependent variable were found to be as $\hat{\rho}_{ML} = 0.569$,

Table 1. Goodness of fit results

Estimation Method for Covariance Parameters	Covariance Structure	-2 Res. Log Likelihood	AIC[1]	BIC[2]
ML	CS, AR(1)	3068.6	3110.6	3153.5
REML		3077.6	3081.6	3085.7
ML	TOEP	3065.7	3109.7	3154.6
REML		3074.1	3080.1	3086.3
ML	UN, CSH, ARH(1),	3008.9	3019.3	3027.5
REML	TOEPH,	3011.3	3054.9	3101.9
	ANTE(1), UNR			

[1]AIC: Akaike Information Criterion; [2] BIC: Bayesian Information Criterion

Table 2. The results of significance of fixed effects

Covariance Structures	Sex		Dam Age		Birth Type		Time 2		Live Weight	
	F*	P*	F	P	F	P	F	P	F	P
CS, AR(1)	4.12	0.0429	2.69	0.0302	19.26	<.0001	30.77	<.0001	764.73	<.0001
	4.16	0.0419	2.63	0.0333	18.99	<.0001	29.91	<.0001	729.24	<.0001
TOEP	4.50	0.0342	2.71	0.0293	19.26	<.0001	32.21	<.0001	744.15	<.0001
	4.56	0.0331	2.66	0.0320	18.99	<.0001	31.38	<.0001	710.89	<.0001
CSH, ARH(1), TOEPH, UN,	1.02	0.3128	1.10	0.3543	5.40	0.0204	43.83	<.0001	571.52	<.0001
ANTE(1), UNR	0.96	0.3272	0.95	0.4353	4.79	0.0290	43.46	<.0001	528.20	<.0001

*: F and P test statistics significant levels

Table 3. Covariance parameter estimates of random effects:

Covariance Structures	Covariance Parameters	Estimation Method for Covariance Parameters	Estimate	Std. Error	Wald Z	Sig.
UN	σ_e^2	ML	3.002	0.175	17.16	<.0001
		REML	3.054	0.179	17.02	<.0001
	σ_1^2	ML	4.072	1.102	3.69	0.0001
		REML	4.785	1.303	3.67	0.0001
	σ_{12}^2	ML	-0.137	0.079	-1.74	0.0818
		REML	-0.137	0.085	-1.62	0.1061
	σ_2^2	ML	0.025	0.008	3.12	0.0009
		REML	0.026	0.008	3.10	0.0010
CSH, ARH(1), TOEPH, ANTE(1),UNR	σ_e^2	ML	3.002	0.175	17.16	<.0001
		REML	3.054	0.179	17.02	<.0001
	σ_1^2	ML	4.072	1.102	3.69	0.0001
		REML	4.785	1.303	3.67	0.0001
	σ_2^2	ML	0.025	0.008	3.12	0.0009
		REML	0.026	0.008	3.10	0.0010
	ρ	ML	-0.427	0.170	-2.51	0.0119
		REML	-0.389	0.177	-2.20	0.0279

$\hat{\rho}_{REML} = 0.605$ for UN covariance model; and as $\hat{\rho}_{ML} = 0.550$, $\hat{\rho}_{REML} = 0.592$ for CSH, ARH(1), TOEPH, ANTE(1), UNR models. UN, CSH, ARH(1), TOEPH, ANTE(1), UNR variance-covariance matrix estimates which were handled in modeling heterogeneity in the data structure from which repeated measures were taken can be shown as follows:

$$\hat{G}_{ML} = \begin{bmatrix} 4.072 & -0.137 \\ -0.137 & 0.025 \end{bmatrix}$$ and

$$\hat{G}_{REML} = \begin{bmatrix} 4.785 & -0.137 \\ -0.137 & 0.026 \end{bmatrix}$$

$$\hat{R}_{ML} = \begin{bmatrix} 3.002 & 0 & 0 & 0 \\ & 3.002 & 0 & 0 \\ & & 3.002 & 0 \\ & & & 3.002 \end{bmatrix}$$ and

$$\hat{R}_{REML} = \begin{bmatrix} 3.054 & 0 & 0 & 0 \\ & 3.054 & 0 & 0 \\ & & 3.054 & 0 \\ & & & 3.054 \end{bmatrix}$$

CONCLUSION

In this study, the RISM model was applied on repeated measure data. Results revealed that UN, CSH, TOEPH, ARH(1), ANTE(1) and UNR heterogeneous variance- covariance models were the best model group in modeling the variance- covariance matrix structure regarding the dependent variable and CS, TOEP and AR(1) homogeneous variance- covariance models were the worst model group. When the heterogeneous variance-covariance models were used as the best model group, the percent of random-effect terms in the model to explain overall change in the dependent variable was 56.9% and 60.5%, this percent declined to 55% and 59.2% when the homogeneous variance- covariance models which were found to be the worst model group were used. CS, TOEP and AR(1) covariance models were thereby found to be the weakest model in explaining the overall change in the structure of the dependent variable. It is quite important to accurately identify the covariance structures between the repeated measures. Because, covariance structures between the repeated measures have an important effect on the estimates to be made.

REFERENCES

Akbaş Y., M. Z. Fırat and Ç. Yakupoğlu (2001). Comparison of different models used in the analysis of repeated measurements in animal science and their sas applications. *Agricultural Information Technology Symposium*. 20-22 September 2001, Sütçü İmam University, Kahramanmaraş.

Antonio, K. and Beirlant, J. (2007). Actuarial statistics with generalized linear mixed models. Mathematics and Economics 40: 58-76.

Doğanay, B. (2007). Mixed Effects Models for Analyzing Longitudinal Studies. Master's Thesis. University of Ankara, Turkey

Eyduran, E. and Akbaş, Y. (2010). Comparison of different covariance structure used for experimental design with repeated measurement. The J. Anim. & Plant Sci., 20(1): 44-51.

Iyit N. (2008). Constitution of Linear Mixed Models in the Analysis of Correlated Data. Ph.D. Thesis. University of Selcuk, Turkey.

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

Orhan. H., E. Eyduran, and Y. Akbaş (2010). Defining the best covariance structure for sequential variation on live weights of anatolian Merinos male lambs. The J. Anim. & Plant Sci. 20(3): 158-163.

Tabachnick B. G. and L. S. Fidel (2001). Using Multivariate Statistics. Allyn & Bacon, USA.