Determination of the Most Appropriate Covariance Structure for Data with Missing Observations in Repeated Measures Design*

Ecevit EYDURAN[1], Adile TATLIYER[2], Abdul WAHEED[3], Mohammad Masood TARIQ[4]**
[1]IU, Faculty of Agriculture, Biometry Genetics Unit, Department of Animal Science, Iğdır-Turkiye
[2]SDU, Biometry Genetics Unit, Faculty of Agriculture, Department of Animal Science, Isparta-Turkiye
[3]Zakariya University, Faculty of Veterinary Sciences, Bahauddin, Multan, Pakistan
[4]University of Balochistan, Center for Advanced Studies in Vaccinology and Biotechnology (CASVAB), Quetta, Balochistan, Pakistan

Received (Geliş Tarihi) **: 12.10.2013** **Accepted** (Kabul Tarihi) **: 16.11.2013**

Abstract: Sequential measurements taken from same experimental unit at different periods of time for a quantitative trait are named Repeated Measurement. In a repeated measures design with two factors (treatment and time), use of MIXED model offers an opportunity for describing various covariance structures (CS, UN, ANTE (1), AR(1), TOEPLITZ etc.) in analyzing data with/without missing observations instead of a Repeated ANOVA (classical approach) in the event of violation of sphericity assumption. In the framework of MIXED modeling used for the repeated measures design, the aims of this study are to evaluate statistical validity of some assumptions relevant to this topic for available data set including missing observations and to get knowledge about selecting the suitable covariance structure for the data set. To achieve these aims, sample data on animal science were provided.

The best covariance structure was selected on the basis of goodness of fit criteria such as AIC, BIC, and AICC. In conclusion, present results obviously reflected that in the case of violation of sphericity assumption, use of MIXED modeling in repeated measures design was a good choice for defining ideal covariance structures for data set with/without missing observations.
Key Words: Covariance Structure, MIXED modeling, Repeated ANOVA, Sphericity Test

Eksik Gözlem İçeren Tekrarlanan Ölçümlü Denemelerde En Uygun Kovaryans Yapısının Belirlenmesi

Özet: Kantitatif bir özellik bakımından aynı deneme ünitesinden farklı zaman periyotlarında alınan ardışık ölçümlere, tekrarlanan ölçüm denir. İki faktörlü (muamele ve zaman) tekrarlanan ölçümlü denemelerde, küresellik varsayımının sağlanmadığı durumda Repeated ANOVA (klasik yaklaşım) yerine MIXED modelinin kullanılması, eksik/tam gözlemli verilerde çeşitli kovaryans yapılarının tanımlanmasına (CS, UN, AR (1), TOEPLITZ vs) olanak sağlamaktadır. Bu çalışmanın amaçları, tekrarlanan ölçümlü denemeler için kullanılan MIXED modelleme kapsamında, eksik gözlemler içeren mevcut veri setleri için konu ile ilgili bazı varsayımların istatistiksel geçerliliğini (doğruluğunu) değerlendirmek ve bu veri setleri için seçilmiş uygun kovaryans yapısının seçilmesi konusunda bilgi sahibi olmaktır. Bu amaçlara ulaşmak için hayvancılık alanında bir veri seti temin edilmiştir. En iyi kovaryans yapısı; AIC, BIC ve AICC uyum iyiliği ölçütlerine göre seçilmiştir.

Sonuçta, bu çalışma, tekrarlanan ölçümlü denemelerde, küresellik varsayımının sağlanmaması durumunda MIXED model kullanımının, eksik/tam gözlem içeren veri setleri için ideal kovaryans yapılarını tanımlamada iyi bir seçenek olduğunu açıkça göstermiştir.
Anahtar Kelimeler: Küresellik testi, Kovaryans yapısı, MIXED model, Repeated ANOVA

INTRODUCTION

Repeated measures are multiple observations taken consecutively on same experimental unit over time for a quantitative (measurable) trait. In the general linear mixed models, repeated measures (correlated) designs that include random and fixed effects (Littell *et al.*, 2000), are likely to be analyzed on the basis of classical (univariate) and advanced (multivariate) approaches (Eyduran and Akbas, 2010). These approaches are often used for assessing repeated measures data with two factors, time and treatment (Orhan et al., 2010). Of them, the simplest one as a univariate approach is Repeated ANOVA, which statistically presents more effectual evaluation in the validity of Spherity assumption in repeated measures data without missing observations. Greenhouse-Geisser and Huynh-Feldt in the violation of this assumption are preferable conventional (univariate) approaches, to a certain extent, compared to Repeated ANOVA. However, Profile analysis (Repeated MANOVA) as a multivariate approach to repeated measures design could produce better valuable results in comparison to these univariate approaches when the assumption is invalid for repeated data with non-missing data. Contrarily to all of these approaches mentioned insofar, applying mixed model approach in repeated measures data is the most

---

*This paper was published as an abstract in VI. Balkan Animal Conference (BALNIMALCON 2013), 3-5 October 2013 in Tekirdag, Turkiye.
**Corresponding author: Eyduran, E., ecevit.eyduran@gmail.com

advantageous optionally, in terms of specifying different covariance patterns with/without missing observations in the violation of Spherity assumption (Eyduran and Akbas, 2010; Orhan *et al.*, 2010).

Over the last decades, some authors have got sophistically in touch with most advantageous of choosing mixed model theory in repeated measures design with two factors (treatment and time) relating to animal husbandry (Littell *et al.*, 1998; Akbas *et al.*, 2001; Pancarci *et al.*, 2007&2009; Orhan *et al.*, 2010; Serbester *et al.*, 2012). Conversely, application of mixed model methodology and determination of ideal covariance structure for repeated measures data with missing observations are scarce in animal science.

The present investigation was undertaken to establish the most appropriate covariance on the basis of applying mixed model theory for repeated measures data with missing observations in animal science. Also, the effect of including covariate on fitting criteria was investigated.

MATERIAL and METHOD

In the current work, 228 Mengali lambs, single-born from dams at the age of 24-30 months, were randomly selected from sheep flocks in Pakistan.

Sex, as a treatment (between –subjects), time as repeated factor (within-subject), and birth weight as a covariate are used throughout this paper. The first 20 observations with the covariate (birth weight) of lambs are shown in Table 1 as also illustrated in Wang and Goonewardane, (2004). All the weight data were taken monthly from birth to six months.

The statistical model used without the covariate was written as Eq.1:

$$y_{ijt} = \text{«} + \sim_i + d_{j(i)} + \text{¥}_t + (\sim\text{¥})_{it} + e_{ijt} \quad (1)$$

Where:

y_{ijt} is the live body weight measured at time t on the j[th] lamb assigned to the i[th] sex, « is the overall mean effect, $\sim_i$ is the i[th] fixed sex effect, $d_{j(i)}$ is the random effect of the j[th] lamb within the i[th]sex $d_{j(i)} \approx NID(0,\pm^2_d)$ (approximately normally independently distributed with mean of 0 and variance $\pm^2_d$), ¥_t is the fixed t[th] time effect when the measurement was taken, $(\sim\text{¥})_{it}$ is the fixed interaction effect between sex and time, e_{ijt} is the random error associated with the jth lamb assigned to the ith sex at the time, t, $e_{ijt} \approx NID(0,\pm^2_e)$ (approximately normally independently distributed with mean of 0 and variance $\pm^2_e$)

Table 1. The first 20 observations of the example data set of Mengali lambs

Sex	ID	Time	BW	Weight
2	1	1	3.50	7.88
2	1	2	3.50	11.73
2	1	3	3.50	15.96
2	1	4	3.50	17.71
2	1	5	3.50	21.43
2	1	6	3.50	25.15
2	2	1	3.50	7.88
2	2	2	3.50	11.73
2	2	3	3.50	15.96
2	2	4	3.50	19.00
2	2	5	3.50	22.99
2	2	6	3.50	26.98
1	3	1	3.60	8.42
1	3	2	3.60	12.63
1	3	3	3.60	17.17
1	3	4	3.60	21.55
1	3	5	3.60	24.79
1	3	6	3.60	28.02
1	4	1	3.90	9.13
1	4	2	3.90	13.68

The statistical model for the mixed model analysis is given Eq. 2

$$y_{ijt} = \text{«} + \sim_i + d_{j(i)} + \text{¥}_t + (\sim\text{¥})_{it} (b + \ddot{}_j) x_{ij} + e_{ijt} \quad (2)$$

Where:

b is the common regression coefficient of initial weight of x_{ij}, $\ddot{}_i$ is the slope deviation of the ith sex from the common slope b, x_{ij} is the initial (birth) weight measure of the lamb j on the sex i at the beginning of the study and the remaining terms are the same as seen in Eq.1.

In this study, the mixed model approach (PROC MIXED) is used with different covariance structures in the SAS system. For example, following SAS commands for TOEPH covariance structure could be written:

```
PROC MIXED;
CLASS SEX ID TIME;
MODEL WEIGHT = SEX TIME SEX*TIME BW
/DDFM=KR;
REPEATED TIME / SUB=ID (SEX) TYPE=
TOEPH;
RUN;
```

DDFM option on the model report in MIXED procedure was used to specify Kenward Roger approach for defining fixed effects and degrees of freedom.

In the repeated measures data, the popular fitting criteria, viz. Akaike Information Criterion (AIC), Burnham-Handerson Criterion (AICC) and Shwartz's

Bayesian Criterion (SBC) were employed in order to find the best covariance structure. For this purpose, repeated (body weight) measurements taken sequentially from 228 Mengali lambs were subjected to different covariance structures: Compound Symmetry (CS), Autoregressive (AR(1)), Unstructure (UN)and Huynh-Feldt (HF), Heterogenous First-Order Autoregressive (ARH(1)), First–Order Ante-Dependence (ANTE(1)), Heterogenous Compound Symmetry (CSH), Toeplitz (TOEP), and Heterogenous Toeplitz (TOEPH).

In the work, applying mixed model methodology in repeated measures data was performed on the basis of three ways:

1. The complete data (referring to all measurements taken completely for each lamb) were exposed to all the covariance structures defined in the above. A covariate for the complete data was not defined.

2. The complete data with a covariate (birth weight) were exposed to various covariance structures mentioned above.

3. Deletion operation was done at different proportions (20% and 40%) of missing observations.

All the statistical evaluations were performed using MIXED procedure of SAS program.

RESULTS and DISCUSSION

In the present work, mixed methodology in repeated measures design was a preferable option, as recommended by many authors (Eyduran and Akbas, 2010; Orhan *et. al.*, 2010), due to the fact that Spherity assumption was violated in the traditional repeated measures design, ignoring the time-dependent correlation (data not shown). In this case, Greenhouse-Geiser (G-G) and Huynh Feldt (H-F) adjustments can be routinely applied in preference to the classical approach, but mixed model approach in repeated measures designs permits a much better specification on selecting the best covariance structure for repeated measures data (Wang and Goonewardene, 2004). Estimates of fitting criteria (-2 Res Log Likelihood, AIC, AICC and BIC) for some covariance structures specified with/without including the covariate in general linear mixed model are given in Table 2. However, with the exception of UN covariance structure, the beneficial impacts of including covariate (birth weight) on specified repeated measures model were determined for the other defined covariance structures. In addition, HF also was the covariance structure that provided the second worst result with/without covariate. The most excellent covariance structure was ANTE (1) structure in the current work with two factors (sex and time) due to the smallest fitting criteria results. Specifying CS, HF, AR (1), and TOEP covariance structure in the repeated measured model including the covariate led to much better results. The worst covariance structure with / without adding the covariate was observed in CS (Table 2).

Many authors reported that the existence of missing measurements was not a significant problem in repeated measures design compared to the classical approach, where it is impossible to accept missing data in repeated measures design. Therefore, acceptability of missing data in the correlated design is an important advantage of mixed model methodology (Eyduran and Akbas, 2010; Orhan et al., 2010).

The significance results of sex, time, and sex by time interaction in the investigated repeated measures data with/without the covariate are presented in Table 3. Including the covariate for most of covariance structures specifying in the studied data gave better results as also confirmed in Table 2.

Significance results for fixed effects in general linear mixed model with/without covariate are summarized in Table 3. It was understood from Table 3 that all the fixed effects, and the covariate in the general linear mixed model with covariate were found significant ($P<0.01$) for other covariance structures, with exception of UN, which did not occur its convergence. In the model without covariate, CS, UN, and TOEPH covariance structure did not provide their convergence, but the other structures ensured convergence operation. In the examination of Table 3, adding the covariate in the built models gave smaller F values.

Table 2. Fitting criteria results for comparing covariance structure with/without a covariate

	With the Covariate				**Without the Covariate**			
Covariance Structure	**-2 Res Log Likelihood**	**AIC**	**AICC**	**BIC**	**-2 Res Log Likelihood**	**AIC**	**AICC**	**BIC**
CS	4667.2	4671.2	4671.2	4678.1	5580.1	5584.1	5584.1	5591.0
CSH	2687.5	2701.5	-	-	-	-	-	-
UN	-	-	-	-	-	-	-	-
HF	3967.1	3981.1	3981.2	4005.1	4620.1	4636.1	4636.2	4663.7
AR(1)	3711.2	3715.2	3715.2	3722.0	4172.6	4176.6	4176.6	4183.5
ARH(1)	2284.4	2298.4	2298.4	2322.4	-		2352.6	2380.0
ANTE(1)	2165.8	2187.8	2188.0	2225.6	2062.0	2088.0	2088.2	2132.7
TOEP	3315.5	3327.5	3327.6	3399.2	3594.9	3608.9	3609.0	3633.0
TOEPH	2263.4	2285.4	2285.6	2323.2	-	-	-	-

Table 3. Significance results of fixed effects in mixed model approach (F and P values)

	With the Covariate				Without the Covariate		
Covariance structure	Sex Effect	Time Effect	Sex by Time Interaction Effect	BW Effect	Sex Effect	Time Effect	Sex by Time InteractionEffect
	F	**F**	**F**	**F**	**F**	**F**	**F**
CS	40.64	6839.56	77.71	140.36	84.94	9512.71	73.25
CSH	84.91	7681.73	30.45	502.53	-	-	-
UN	-	-	-	-	-	-	-
HF	63.88	6853.40	77.53	22.89	86.62	9529.38	73.13
AR(1)	21.70	4398.40	50.12	67.52	42.25	5562.95	45.21
ARH(1)	84.59	5648.36	31.29	268.34	97.42	6511.43	27.40
ANTE(1)	85.91	6751.28	36.35	303.53	83.05	6020.91	30.58
TOEP	18.11	3441.80	53.17	53.77	32.88	2480.32	29.93
TOEPH	86.14	5548.57	32.74	280.75	-	-	-

All F values estimated were very significant (P<0.001)

For the repeated measures data of the missing observations, a specification of mixed model methodology together with a variety of covariance structures in repeated measures design offers the most important advantages, whereas the analysis of the repeated measures data containing missing observations for Repeated ANOVA and Profile Analysis is assuredly unacceptable (Eyduran and Akbas, 2010).

In the current work, in order to investigate the effect of various missing observation proportions (20% and 40%) and to determine the suitable covariance structure for the repeated measures data of the missing observations, some observations at 20% (one observation) and 40% (two observations) proportions from each animal were deleted randomly in the application of mixed model methodology. Deletion operation was made on the basis of other sequential measurements with the exception of birth weight (covariate). Hence, the fitting criteria results and F values found for several covariance structures specified for the current work are summarized in Tables 4 and 5, respectively. ANTE (1) was defined to be the best covariance structure for both missing proportions. ANTE (1) covariance structure indicates this kind of covariance structure (unequal variances over time and unequal correlations among different pairs of measurements). However, CS was the worst covariance structure that gave the greatest fitting criteria values. For instance, convergence operation was stopped for HF and UN at 20% missing data proportion (Table 4).

In the examination of the Table 5, sex, time, sex by time interaction and covariate (birth weight-BW) effects were very significant in using mixed model for repeated measures data of missing observations (P<0.001), with the exception of HF at %20 missing proportion and UN at both %20 and 40% missing data proportions.

Table 4. Fitting criteria results for comparing covariance structures at different missing proportions in the model with adding the covariate

	20(%) missing observation				40(%) missing observation			
Covariance Structure	-2 Res Log Likelihood	AIC (smaller is better)	AICC (smaller is better)	BIC (smaller is better)	-2 Res Log Likelihood	AIC (smaller is better)	AICC (smaller is better)	BIC (smaller is better)
CS	3977.2	3981.2	3981.2	3988.0	3194.9	3198.9	3198.9	3205.8
CSH	2270.5	2284.5	2284.6	2308.6	1800.2	1814.2	1814.3	1838.2
UN	-	-	-	-	-	-	-	-
HF	-	-	-	-	2616.8	2630.8	2630.9	2654.8
AR(1)	3311.5	3315.5	3315.5	3322.4	2812.3	2816.3	2816.3	2823.2
ARH(1)	1940.8	1954.8	1954.9	1978.9	1624.3	1638.3	1638.4	1662.3
ANTE(1)	1871.7	1893.7	1893.9	1931.5	1550.3	1572.3	1572.6	1610.1
TOEP	3042.3	3054.3	3054.4	3074.9	2698.2	2710.2	2710.3	2730.8
TOEPH	1901.8	1923.8	1924.0	1961.6	1611.5	1633.5	1633.8	1671.3

Table 5. Significance results of fixed effects in mixed model approach (F and P values)

	20(%) missing observation				40(%) missing observation			
	Sex Effect	Time Effect	Sex by Time Interaction Effect	BW Effect	Sex Effect	Time Effect	Sex by Time Interaction Effect	BW Effect
Covariance structure	F	F	F	F	F	F	F	F
CS	39.39	5640.01	61.06	137.07	41.99	4552.24	47.36	138.77
CSH	76.70	7098.85	26.61	539.66	78.13	6072.06	24.27	435.96
UN	-	-	-	-	-	-	-	-
HF	-	-	-	-	72.17	4791.13	47.23	290.24
AR(1)	22.71	4043.97	41.51	73.86	25.83	3653.78	36.88	90.83
ARH(1)	79.36	6141.81	24.87	294.09	79.66	5926.65	24.42	284.95
ANTE(1)	85.23	7367.73	33.25	254.34	85.97	7065.54	32.78	222.91
TOEP	22.69	2554.83	26.44	76.53	26.01	2596.51	27.11	83.86
TOEPH	82.63	5805.92	25.64	308.42	80.20	5923.65	25.14	286.95

All F values estimated were very significant ($P<0.001$)

Univariate ANOVA, Greenhouse-Geisser Epsilon (G-G) adjusted F test, Huynh-Feldt Epsilon (H-F) adjusted F test and profile analysis (Repeated MANOVA) could not use for repeated measures data with missing observations (Littell *et al*., 1996; 1998; Akbaş *et al*., 2001). Because of the fact that mixed model methodology has much more advantageous and more reliable results, many authors have reported that the application of mixed model methodology was a better choice for repeated measures data with/without missing data (Littell et al., 1998; Eyduran and Akbaş, 2010; Orhan et al., 2010).

CONCLUSION

In repeated measures data, the univariate approach is the best choice in the validity of Spherity assumption without missing data. However, in the event of violation of sphericiticity assumption, mixed model methodology was much more suitable application than Univariate ANOVA, Greenhouse-Geisser Epsilon (G-G) adjusted F test, and Huynh-Feldt Epsilon (H-F) adjusted F test and Repeated MANOVA without missing data. As mentioned earlier, applying mixed model methodology in the violation of the assumption as also observed in the current work was the best selection to specify various covariance structures in the repeated measures data with/without missing observations. The current work gave some important results in the context of mixed model methodology in repeated measures data that:

1. Adding covariate in the specified models for CS, HF, and AR (1) provided much better contributions to the improvement of fitting criteria results.
2. ANTE (1) covariance structure was the best selection procedure for the repeated measures data without missing observations.
3. ANTE (1) covariance structure was determined to be the most appropriate procedure the data sets at different proportions (20% and %40) of missing observations.
4. The covariate effect (birth weight) was found very significant ($P<0.001$) for repeated measures data with/without missing data.
5. Sex, time, and sex by time interaction fixed effects were very significant ($P<0.001$).

In conclusion, applying mixed model methodology is recommended in repeated measures data of missing observations in contrast to classical approaches.

REFERENCES

Akbaş, Y., Fırat, M.Z., Yakupoğlu, Ç. 2001. Comparison of Different Models Used in the Analysis of Repeated Measurements in Animal Science and Their SAS Applications. Agricultural Information Technology Symposium, Sütçü İmam University, Agricultural Faculty, Kahramanmaraş, 20-22 September 2001.

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

Littell, R. C., Henry, R. C., Ammerman, C. B. 1998. Statistical analysis of repeated measures data using SAS procedures. J. Anim. Sci. 76: 1216–1231.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. 1996. SAS system for mixed models. SAS Institute, Inc., Cary, NC.

Littell, R. C., Pendergast, J., Natarajan, R. 2000. Modeling covariance structure in the analysis of repeated measures data. Stat. Med. 19: 1793–1819.

Orhan H., Eyduran, E., Akbaş, Y. 2010. Defining The Best Covariance Structure for Sequential Variation on Live Weights of Aanatolian Merinos Male Lambs. J. Anim. and Plant Sci., 20(3), 158-163.

Pancarci S.M., Kacar, C., Ogun, M., Gungor O., Gurbulak, K., Oral, H., Karapehlivan, M., Citil M. 2007. Effect of L-Carnitine Administration on Energy Metabolism During Peripaturient Periodin Ewes. Kafkas Univ. J. Vet. Faculty. 13(2): 149-154.

Pancarci, S.M., Gurbulak, K., Oral, H., Karapehlivan, M., Tunca, R., Çolak A. 2009. Effect of Immunomodulatory Treatment with Levamisole on Uterine Inflammation and Involution, Serum Sialic Acid Levels and Ovarian Function in Cows. Kafkas Univ. J. Vet. Faculty. 15(1): 25- 33.

Serbester, U., Çınar, M., Ceyhan, A., Erdem, H., Görgülü, M., Kutlu, H.R., Baykal Çelik, L., Yücelt, Ö., Cardozo, P.W. 2012. Effect of Essential Oil Combination on Performance, Milk Composition, Blood Parameters and Pregnancy Rate in Early Lactating Dairy Cows during Heat Exposure. J. Anim. and Plant Sci., 22(3): 556-563, ISSN: 1018-7081.

Wang, Z., Goonewardene, L. A. 2004. The use of MIXED models in the analysis of animal experiments with repeated measures data. Can. J. Anim. Sci., 84(1):1-11.