Asian Journal of Agricultural Extension, Economics & Sociology

40(10): 33-40, 2022; Article no.AJAEES.86339
ISSN: 2320-7027

An Extension of the Mixed Effects Model that Includes Correlation within Group Errors

Nageena Nazir [a*], Shakeel Ahmad Mir [a], Zahoor Ahmed Rather [b] and Masroor Ahmad Dar [a]

[a] Division of Agricultural Statistics, Sher-e-Kashmir University of Agricultural Sciences and Technology, Shalimar, Kashmir, India.
[b] Division of Floriculture and Landscaping, Sher-e-Kashmir University of Agricultural Sciences and Technology, Shalimar, Kashmir, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJAEES/2022/v40i1031035

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/86339

Original Research Article

Received 18 February 2022
Accepted 23 April 2022
Published 08 July 2022

ABSTRACT

Mixed effect models allow for a great deal of flexibility in defining random outcomes, but they limit within-group errors to independent disbursed random variables with a zero mean and constant variance. In addition to its random outcomes for the implied structure, this model is expanded in this paper by including within-group correlated errors. We demonstrate how to accurately predict the model's parameters using a marginal maximum probability (ML) method. The model's accuracy is demonstrated by a real-world example. Additionally, we provide several instructions for correlation systems to represent serial and spatial correlation. Finally, we define how to combine variance features and correlation systems to flexibly model the within-group variance-covariance shape. We also discuss how the lme function can be used to maintain the prolonged linear mixed effects model. And, when compared to other models, the exponential spatial correlation version has the smallest AIC and BIC, making it appear to be the best within-group correlation model.

Keywords: Linear mixed model; correlated; correlation structures.

**Corresponding author: E-mail: nazir.nageena@gmail.com;*

1. INTRODUCTION

For a variety of grouped information types that were discovered during the exercise, these basic linear mixed effects model provides a good model. Although many applications involving grouped records have inside group errors that are heteroscedastic (i.e., have unequal variances), correlated, or both, this is not always the case [1-5]. In order to accommodate heteroscedastic within-group errors, we volume the primary linear combined results version in this paper. Many authors have suggested such extensions to the mixed-outcomes version, with Hedeker et al.'s combined-outcomes area scale model being the most well-known (but see also Lee and Nelder [6], Gasimova et al. [7], Hamaker et al. [8], Schuurman et al. [9], Wang et al. [10], and Nordgren et al. [11].

2. MATERIALS AND METHODS

This article's main goal is to describe an extension of this model that also includes correlated within-group errors and is conceptually similar to the random effects for the mean shape. Furthermore, we demonstrate how a marginal ML method can be used to successfully predict the parameters of the model. The MEM is first introduced in the following, along with the aforementioned extensions. We have used an actual animal technology statistics set to validate the extended model. For the current study, the records produced from the test on chicks conducted with assistance from the division of LPT SKUAST-Kashmir, Shuhama, were used. Statistics on the body weight of chicks that were measured over a period of 64 days were compiled. On day 1 and then every 7 days after that, the body weight changed, and on day 44, a new dimension was added. As a result, each chicken provided an average of eleven readings. There were 3 groups of chicks on unique diets, with 5 chicks in every group. For analysis and modeling in the R/SAS program, the data set was given the name Chick Weight. Four columns and 165 rows make up its structure.

The column headings for the body weight of chicks, the time interval, the number of chicks, and the diet are weight, time, chick, and diet, respectively.

2.1 Preferred Extended Linear Mixed Effects Model Formulation

We assume heterocedastic and correlated within-group errors in the extended single-level linear mixed effect model, and these errors can be expressed as follows:

$$y_i = x_i\beta + z_i b_i + e_i \quad (2.1)$$
$$b_i \sim N(0, \Psi),\ e_i \sim N(0, \sigma^2\Lambda_i)$$

where the Λ_i is positive particular matrices parameterized via a fixed typically small, set of parameters λ.

2.2 Strategies for Computation and Estimation

For linear mixed effects models, a variety of parameter estimation techniques have been employed; the same will be done for extended linear mixed effects models. The two generic techniques among them are ML and REML. The specific description of the two can be found in [12], section 2.2). Because Λ_i is a positive definite, it admits an invertible square root $\Lambda_i^{½}$ [13] with an inverse $\Lambda_i^{-½}$ such that $\Lambda_i = (\Lambda_i^{½})^T \Lambda_i^{½}$ and $\Lambda_i^{-1} = \Lambda_i^{-½}(\Lambda_i^{-½})^T$.

Letting $y_i^* = (\Lambda_i^{-½})^T y_i$, $e_i^* = (\Lambda_i^{-½})^T e_i$ $x_i^* = (\Lambda_i^{-½})^T x_i$, $z_i^* = (\Lambda_i^{-½})^T z_i$ (2.2)
and noting that $e_i^* \sim N[(\Lambda_i^{-½})^T 0,\ \sigma^2 (\Lambda_i^{-½})^T \Lambda_i \Lambda_i^{-½}] = N(0, \sigma^2 I)$.

Thus the model 2.1 can be revived as

$$y_i^* = x_i^*\beta + z_i^* b_i + e_i^* \quad (2.3)$$
$$b_i \sim N(0, \Psi),\qquad e_i^* \sim N(0, \sigma^2 I)$$

The extended linear mixed effects model's likelihood function, L, is represented by the following equation.

$$L(\beta,\theta,\sigma^2,\lambda \mid y) = \prod_{i=1}^{M} p(y_i \mid \beta,\theta,\sigma^2,\lambda) \quad (2.4)$$

We must combine the conditional density of the data given the random effect in order to obtain the marginal density for the data because the non-observable random vector bi i = 1,...M is a component of the model. To express this, we will use the independence of bi and ei as follows:

$$\prod_{i=1}^{M} p(y_i^* \mid \beta,\theta,\sigma^2,\lambda) \mid \Lambda_i^{-1/2} = L(\beta,\theta,\sigma^2,\lambda \mid y^*) \prod_{i=1}^{M} \Lambda_i^{-1/2} \quad (2.5)$$

The conditional density of yi and the marginal density of bi are both multivariate normal, where

P (.) denotes a probability density function. Due to the fact that the likelihood L (β, , , λ|y*) corresponds to a fundamental linear mixed effects model, all of the LMEM's effects (see Pin Herio and Bates2000 section 2.2) properly apply to this.

As a result, the likelihood can be evaluated using an orthogonal triangular decomposition just as in the case of a simple linear mixed effects model, leading to a numerically efficient algorithm for maximum likelihood estimation.

Although in theory the random effects biaren't statistical model parameters, they behave in some ways like parameters and we frequently need to estimate their values. The best linear unbiased predictors, also known as BLUP's, of the bi, i = 1, 2, ---M are the conditional models of the random effects evaluated at the conditional estimate of β. The matrices produced by the orthogonal triangular decomposition can be used to evaluate them. In actual use, the maximum likelihood estimate of the unknown vector is used to produce estimated BLUPs.

2.3 Decomposing the within-group Variance Covariance Structure

Usually, the Λ_i matrices can be broken down into a constructed form of simpler matrices:

$\Lambda_i = V_i C_i V_i$, It is simple to verify that: var (e_{ij}) = $\sigma^2 [v]_{ij}^2$, and $cor(e_{ij}, e_{jk}) = [c_i]_{jk}$

In order for C_i to describe the correlation of the within group errors e_i and V_i to describe that variance, it is practical from a theoretical and computational standpoint to divide Λ_i into variance structure elements and a correlation structure component. It enables us to individually model, create codes for, and integrate the two structures into a family of adaptable models for the within-group variance covariance.

2.4 Correlation Structure for Modelling Dependence

Modeling the interdependence of observations is done using correlation structures. They are used to model the dependencies between the errors within groups in mixed effects models [14-20]. Historically, the two main classes of data for which correlation structures had been developed were time series data and spatial data. We anticipate that the within-group errors eij are linked to the position vector pij in order to establish a general framework for correlation structures. This study makes the assumption that the correlation structures are isotropic [13]. It is possible to write the general within group correlation structure for single level grouping for i = 1, --- M and = 1,...,ni as follows.

Cor (e_{ij}, e_{ij}') = h [d (p_{ij}, p_{ij}'), ρ]

where h(.) is a correlation function with values between -1 and 1, and is a vector of correlation parameters.

2.5 Spatial Correlation Structures

"These were initially put forth to model dependence in data, such as geostatistical data, lattice information, and point styles, indexed by continuous two-dimensional position vectors. Here, we only take into account isotropic spatial correlation structures, which are easily generalized to any finite number of position dimensions and can be expressed as continuous functions of the same distance between position vectors. In the mixed effects models, Diggle et al. (1994) serve as a straightforward reference for the spatial correlation shape" (Cressie, 1993).

Spatial correlation structures can typically be represented by their semivariogram, n preference to their correlation feature (Cressie, 1993). The semivariogram of an isotropic spatial correlation structure with a distance function (d). can be defined as :

The classical estimator of the semivariogram (Matheron, 1962) may be expressed as

In which N(s) denotes the quantity of residual pairs at a distance of each other.

2.6 Correlation Structures in Nlme

The nlme library of R-software program presents a hard and fast set of instructions for correlation systems, the corstruct classes, which can be used to specify within-group correlation models in the extended linear mixed effects. Value and form are the two main arguments used by the majority of constructors [21-24]. Correlation structures can be specified as corstruct objects, created using the standard constructor. In this section, we'll examine some examples of grouped data with correlated within-group errors in order to explain how correlation models are used in lme(). Miles is frequently helpful to recall and diagnose plots of the normalised residuals while evaluating the suitability of a correlation model. The normalised residuals should roughly follow an independent N (0, 1) random vector distribution if the within-group variance covariance model is accurate. The Body Weight fact set, as mentioned in the introduction, is used to demonstrate the usage of corStruct classes in lme in combination with variance functions [25-27]. To fit continuous time within-group correlation models, which clearly account for the data imbalance, we employ the spatial correlation cor Struct() classes. Cor Exp, corGaus, corLin, corRatio, and corSpher are the corStruct classes that represent spatial correlation structures. The range of the argument value is utilised to determine An optional grouping variable and a position vector are specified in the argument form, which is a one-sided formula. The position vector's coordinates can be anything, but they must be numerical variables. A character string called metric is used as the parameter to specify the metric to be applied for determining the separations between pairs of distances.

The Variogram approach for the lme() class estimates the sample semivariogram from the residuals of the lme() object. The arguments resType and robust control, respectively, are used to decide what type of residuals should be used ("pearson" or "response") and whether the robust algorithm () or the classical algorithm () should be used to estimate the semivariogram. The defaults are resType = "pearson" and robust = FALSE,in order that classical estimates of the semivariogram are obtained from the standardized residuals. The argument form is a one-sided system specifying the position vector for use for the semivariogram calculations. The effects obtained by using the varCorr model are summarised in Table 1.

The semivariogram, the space, and the number of residual pairs used in the estimation are each represented by a column inside the Tabl-1 returned by the variogram. Since the number of residual pairs used at each distance varies significantly due to the imbalance in the time measurements, some estimates of the semivariogram are more trustworthy than others [28-30]. The wide range of residual pairs that are utilized in the semivariogram estimation generally decreases with distance, rendering the values at great distances unreliable. The argument max Dist() will be used to control the maximum distance for which semivariogram estimates should be computed. The sample semivariogram is depicted graphically in Fig. 1 using the plot method for class semivariogram.

A loess smoother will improve the visualization of semi-variogram patterns, as seen in Fig. 1. Up to 20 days later, the semivariogram seems to bloom with increasing distance before stabilizing at 1. Because of this, we'll use an exponential spatial correlation model to account for within-group errors. Table 2 summarizes the results.

Table 1. Variograms of the data obtained at different distances for different pairs of observations

Variogram	Distance	No. of pairs
0.34	1	16
0.99	6	16
0.76	7	144
0.68	8	16
0.68	13	16
0.95	14	128
0.89	15	16
1.69	20	16
1.12	21	112
1.08	22	16
0.89	28	96
0.93	29	16
0.85	35	80
0.75	36	16
1.08	42	64
1.56	43	16
0.64	49	48
0.67	56	32
0.58	63	16

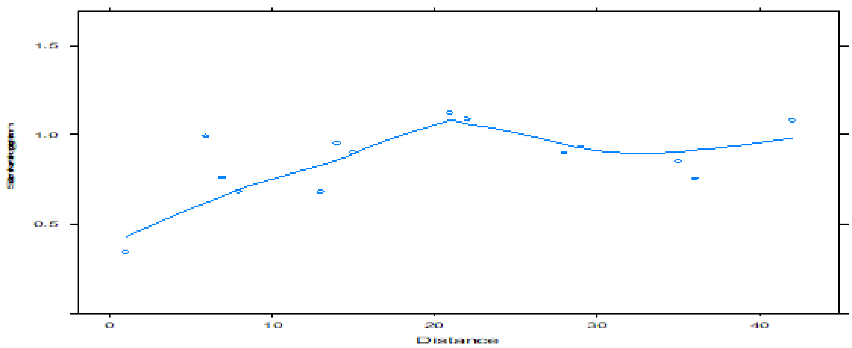

Fig. 1. Pattern semi-variogram estimates for the fitted varPower model's standardized residuals. A loess smoother is added to enhance the visualization of semi-variogram patterns

Table 2. A summary of the fixed effects results from the exponential spatial correlation model

	Value	Standard Error	t value	p value
(Intercept)	251.65	13.06	19.25	0.0000
Time	0.36	0.088	4.08	0.0001
Diet2	200.78	22.65	8.86	0.0000
Diet3	252.58	22.66	11.12	0.0000
Time:Diet2	0.620	0.16	3.87	0.0002
Time:Diet3	0.3	0.15	1.89	0.0601

Table 3. 95% confidence intervals for fixed effects

	Estimate	Lower	Upper
Intercept	251.48	225.63	277.33
Time	0.36	0.19	0.54
Diet2	200.78	151.72	249.84
Diet3	252.57	203.51	301.66
Time:Diet2	0.06	0.032	0.09
Time:Diet3	0.31	0.003	0.61

Table 4. 95% confidence intervals for random effects

	Estimate	Lower	Upper
Sd(intercept)	36.91	25.03	54.45
sd(Time)	0.23	0.15	0.37
Cor((Inter.),Time)	-0.15	-0.62	0.41
Variance function Power	0.59	0.25	0.94
Correlation Structure range	4.88	2.46	9.69
Within Group Standard error	0.14	0.02	1.05

The values of various fixed effects and their interaction values, along with their standard error, t-cal and p-value, are provided in Table 2. Take note that the exponential spatial correlation model has a variance function as well as a correlation shape. Using the intervals approach, we estimate the range of the spatial correlation parameter.

Tables 3 and 4 show the approximate 95% confidence intervals for the fixed and random effects of the Exponential Spatial Correlation model.

The fact that the confidence intervals in each table are bounded away from zero suggests that the spatial correlation model produced a noticeably better fit.

We can put this to the test using the anova method. Table 5 displays the outcomes from the anova technique.

Table 5. Empirical comparison of the fitted models such as the exponential spatial correlation model and the varPower (heteroscedastic) mixed effects model

Models	AIC	BIC	Log Lik	Likelihood Ratio Test	p value
Var Power Model	1163.92	1198.41	-570.96		
Exponential Spatial correlation Model	1145.14	1182.77	-560.57	20.78	<.0001

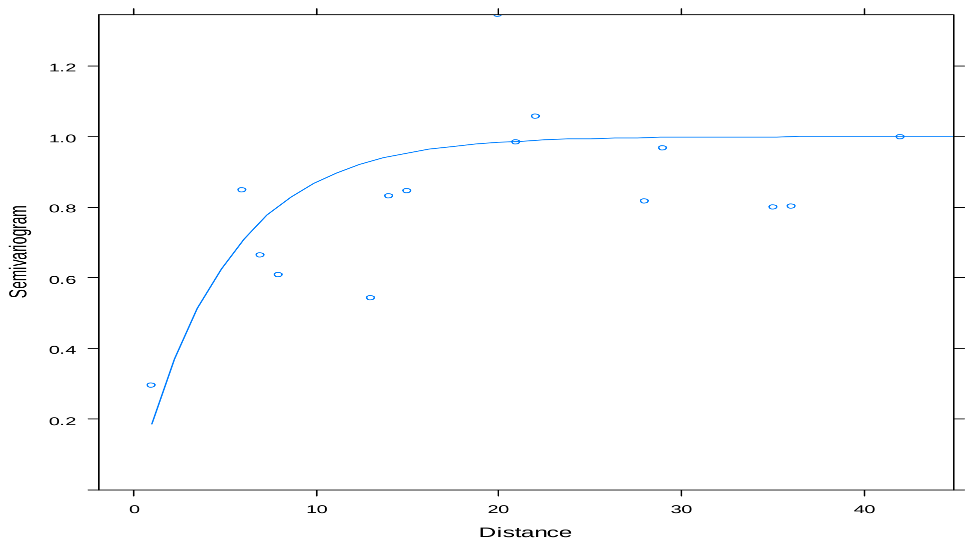

Fig. 2. Sample semivariogram estimates corresponding to the standardized residuals of the fitted varPower model Plot is updated with the fitted semivariogram for the exponential spatial correlation model

Table 5 shows that the exponential spatial correlation version has the lowest AIC/BIC value. The exponential spatial correlation model therefore performs better than the varPower model because the model is more accurately fitted when the AIC/BIC value is lower. The likelihood ratio test additionally indicates that the corExp model fits the data notably higher than the independent errors model corresponding to varPower model.

We can also assess the suitability of the correlation version with the plotting technique when an LME object also contains a spatial corStruct object. Instead of a Loess smoother in this case, the fitted semivariogram for the corStruct object is shown in the plot along with the sample variogram estimates.

The fitted semivariogram, which corresponds to the exponential spatial correlation version, is added to the plot, and it agrees reasonably well with the sample variogram estimates. Examining the sample semivariogram for the normalized residuals allows us to judge the efficacy of the exponential spatial correlation model.

To make it easier to see patterns in the semivariogram, a loess smoother is added to the plot. The effect of an outlying value at distance 1 on the loess smoother is minimized using the robust semivariogram estimator. The sample semivariogram estimates in Fig. 3 seem to fluctuate haphazardly around the y = 1 line, indicating that the normalized residuals are roughly uncorrelated and that the corExp model is suitable.

The update method and the anova method can be used to contrast the exponential spatial correlation model with other spatial correlation models. Since the models are not nested, we can evaluate them using the AIC and BIC information criterion. The results from the anova method are summarized in Table 6 below.

Table 6 shows that the Exponential spatial correlation version has the smallest AIC and BIC when compared to other models, making it the most suitable within-group correlation model for the Bodyweight data. The same is depicted from the log likelihood values.

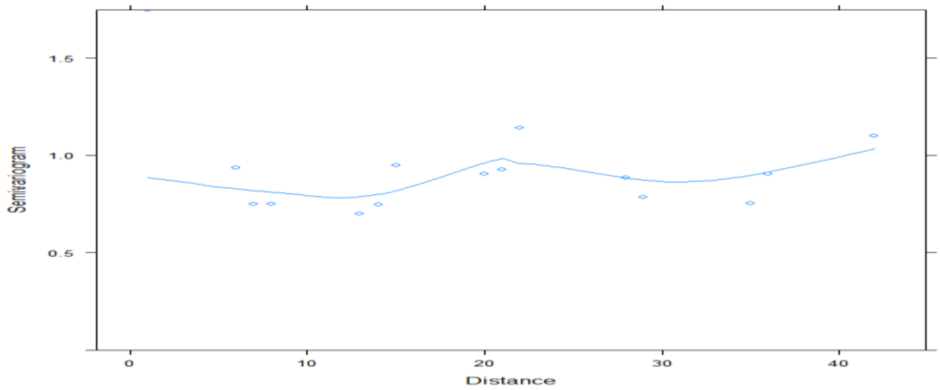

Fig. 3. Pattern semi-variogram estimates corresponding to the normalized residuals of the fitted exponential spatial correlation model

Table 6. Empirical comparison of the various Spatial correlation models

Models	AIC	BIC	logLik
Exponential spatial cor	1145.14	1182.77	-560.57
Ratio spatial correlation	1148.76	1186.39	-562.57
Sphere spatial correlation	1150.78	1188.41	-563.39
Linear spatial correlation	1150.78	1188.41	-563.39
Gaussian spatial correlation	1150.78	1188.41	-563.39

3. CONCLUSION

Therefore, it may be concluded that the Exponential spatial correlation model, among the various models used to extend the basic linear mixed effect model to include the inside the within group correlated errors, has the smallest AIC and BIC compared to other models and thus seems to be the most maximum acceptable within-group correlation model.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Longford NT. Random Coefficient Models, Oxford University Press, New York; 1993.
2. Nestler S. An extension of the mixed-effects growth model that considers between-person differences in the within-subject variance and the autocorrelation. Statistics in Medicine. 2022;41(3):471-482. DOI: 10.1002/sim.9280
3. Nestler S. An extension of the mixed-effects growth model that considers between-person differences in the within-subject variance and the autocorrelation. Statistics in Medicine. 2022;41(3):471-482. DOI: 10.1002/sim.928
4. Nestler S. An extension of the mixed-effects growth model that considers between-person differences in the within-subject variance and the autocorrelation. Statistics in Medicine. 2022;41(3):471-482. DOI: 10.1002/sim.9280
5. Nestler S. an extension of the mixed-effects growth model that considers between-person differences in the within-subject variance and the autocorrelation. Statistics in Medicine. 2022;41(3):471-482. DOI: 10.1002/sim.9280
6. Lee Y, Nelder J A. Double hierarchical generalized linear models (with discussion). J Royal Stat Soc Ser C (Appl Stat). 2006;55:139-185.
7. Gasimova F, Robitzsch A, Wilhelm O, Hülür G. A hierarchical Bayesian model with correlated residuals for investigating stability and change in intensive longitudinal data settings. Methodology. 2014;10:126-137.
8. Hamaker EL, Asparouhov T, Brose A, Schmiedek F, Muthén B. At the frontiers of modeling intensive longitudinal data: dynamic structural equation models for the affective measurements from the COGITO

study. Multivar Behav Res. 2018;53:820-841.15.
9. Schuurman NK, Grasman RPPP, Hamaker EL. A comparison of inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models. Multivar Behav Res. 2016;51:185-206.
10. Wang LP, Bergeman CS, Hamaker E. Investigating inter-individual differences in short-term intra- individual variability. Psychol Methods. 2012;17:567-581.
11. Nordgren R, Hedeker D, Dunton G, Yang CH. Extending the mixed-effects model to consider within-subject variance for ecologicalmo mentary assessment data. Statistics in Medicine. 2020; 39:577-590.
12. Thisted RA. Elements of Statistical Computing, Chapman & Hall, London; 1988.
13. Pinheiro JC, Bates MD. Mixed-Effects Models in S and S PLUS. Springer-Verlag New York; 2000.
14. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Ser B. 1977;39:1-22.
15. Gelman A, Pardoe I. Bayesian measures of explained variance and pooling in multilevel (hierarchical) models Technometrics. 2006;48(2):241-251.
16. Hedeker D, Mermelstein RJ, Berbaum ML, Campbell RT. Modeling mood variation associated with smoking: an application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. Addiction. 2009; 104:297-307.
17. Hedeker D, Mermelstein RJ, Demirtas H. An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. Biometrics. 2008;64:627-634
18. Hedeker D, Demirtas H, Mermelstein RJ. A mixed ordinal location scale model for analysis of ecological momentary assessment (EMA) data.Stat Interf. 2009; 2:391-401.9.
19. Jiang J. Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. Statistica Sinica.1998;8(3):861-886
20. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982;38:963-974.
21. Bates D, Maechler M, Bolker B. mlmRev: Examples from Multilevel Modelling Software Review. R package version. 2011;1:0-1.
22. Bates D, Maechler M, and Bolker B. MEMSS: Data sets from Mixed Effects Models in S. R package version. 2012;0:9-0.
23. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed effects models using lme4. Journal of Statistical Software. 2015; 67:1-48.
24. Davidian M, Giltinan DM. Nonlinear Models for Repeated Measurement Data, Chapman & Hall, London; 1995.
25. Lindstrom MJ, Bates DM. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. Journal of the American Statistical Association.1988;83:1014-1022.
26. Loy A. Diagnostic tools for hierarchical (multilevel) linear models. R package version. 2013; 0.2.2.
27. Nestle S. Extension of the mixed-effects growth model that considers between-person differences in the within-subject variance and the autocorrelation. Statistics in Medicine. 2022; 41(3):471-482. DOI: 10.1002/sim.9280
28. Pinho LGB, Nobre SJ, Freilas SM. On linear mixed models and their influence diagnostics applied to an actuarial problem. Chilean Journal of Statistics. 2012;3(1):57-73.
29. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2013.
30. Wickham H. ggplot2: Elegant graphics for data analysis. Springer New York; 2009.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/86339