

Article

Repeated-Measures Analysis in the Context of Heteroscedastic Error Terms with Factors Having Both Fixed and Random Levels

Lyson Chaka *  and Peter Njuho

Department of Statistics, University of South Africa, Johannesburg 1709, South Africa; njuhopm@unisa.ac.za

* Correspondence: lyson.chaka@spu.ac.za

Abstract: The design and analysis of experiments which involve factors each consisting of both fixed and random levels fit into linear mixed models. The assumed linear mixed-model design matrix takes either a full-rank or less-than-full-rank form. The complexity of the data structures of such experiments falls in the model-selection and parameter-estimation process. The fundamental consideration in the estimation process of linear models is the special case in which elements of the error vector are assumed equal and uncorrelated. However, different assumptions on the structure of the variance–covariance matrix of error vector in the estimation of parameters of a linear mixed model may be considered. We conceptualise a repeated-measures design with multiple between-subjects factors, in which each of these factors has both fixed and random levels. We focus on the construction of linear mixed-effects models, the estimation of variance components, and hypothesis testing in which the default covariance structure of homoscedastic error terms is not appropriate. We illustrate the proposed approach using longitudinal data fitted to a three-factor linear mixed-effects model. The novelty of this approach lies in the exploration of the fixed and random levels of the same factor and in the subsequent interaction effects of the fixed levels. In addition, we assess the differences between levels of the same factor and determine the proportion of the total variation accounted for by the random levels of the same factor.



Citation: Chaka, L.; Njuho, P. Repeated-Measures Analysis in the Context of Heteroscedastic Error Terms with Factors Having Both Fixed and Random Levels. *Stats* **2022**, *5*, 458–476. <https://doi.org/10.3390/stats5020027>

Academic Editor: Wei Zhu

Received: 12 March 2022

Accepted: 15 April 2022

Published: 6 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: covariance structure; linear mixed models; repeated measures; within-subject factor; sphericity; compound symmetry; ANOVA

1. Introduction

Linear mixed-effects models can handle both fixed and random effects simultaneously. Not only are the models convenient for modelling the means of the data, but they are also convenient for modelling covariances [1,2]. Situations arise in fields such as psychology and medicine in which longitudinal or correlated data display notorious heterogeneity of responses to stimuli and treatment [3]. When the classical linear mixed-effects model is inappropriate, alternative approaches, which identify and consider subgroups in fixed and random effects, are necessary. The authors of [4] propose a heterogeneity model as an extension to that proposed by the authors of [5,6] by replacing the normality assumption for random effects and by assuming a non-zero mean vector and a common error vector.

Repeated-measures data are often dependent, a property that does not conform to the generality of a mixed-effects model [1,7,8]. Linear mixed-effects models are some of the most convenient statistical approaches, which account for this dependency [9]. However, setting them up for data analysis requires some care, especially in choosing the most appropriate covariance structure to keep the type I error down [10]. Selecting a suitable model for the covariance is important because the precision of the confidence intervals and the tests of hypotheses concerning model parameters depend on the correct model [11]. According to [12], selecting an appropriate mixed-effects model and construction approach, such as partitioning of fixed and random effects, allows for analysis of variance in correlated data. Numerous diagonal and non-diagonal covariance structures for correlated data,

which cover a range of assumptions about the associations between responses from the same cluster data, are available [13,14]. Modern statistical software, such as SAS Studio in Linux Environment, provides options for selecting candidate covariance structures through the *PROC MIXED* procedure [15] (see Appendix A). The residual maximum likelihood (REML) is one of the most famous methods for estimating covariance parameters associated with linear mixed models [16,17], among other alternatives [18].

We present the construction and analysis of a three-factor linear mixed model for repeated-measures designs when the between-subjects factors consist of both fixed and random levels, and the structure of the variance–covariance matrix of the error terms is not the identity. We consider a repeated-measures design setting under a linear mixed-effects model, with factors sharing both fixed and random components of the model. For experimental designs consisting of factors with a unique composition of both fixed and random levels, we propose a partitioning approach based on factor levels in model construction, estimation, and hypothesis testing. In such a situation, the fixed levels allow for the comparison of specific levels of interest within the factor, whereas the random levels allow for the assessment of variation within the same factor. We assess the effect of introducing heterogeneity of error terms in the selection of the most appropriate covariance structure, in the assessment of the changes that occur in the estimation, and when drawing inferences. To simplify the proposed approach, we focus on the diagonal covariance structure for the repeated-measures linear mixed-effects model.

Section 2 presents the approach to model construction, for a general linear mixed-effects model in a completely randomised design (CRD) and in a repeated-measures design (RMD) under the default assumption of error terms. Section 3 presents results from a numerical example. Discussion of the results is in Section 4, followed by the conclusions in Section 5.

2. Materials and Methods

2.1. An Illustrative Data Structure

We motivate the approach using data collected from a study that investigated the impact of combining carbon tetrachloride (CCl_4) with four levels (0, 1.0, 2.5, and 5.0 mM) and chloroform ($CHCl_3$) with four levels (0, 5, 10, and 25 mM) on the toxicity of cells on in vitro toxicity of isolated hepatocyte suspensions [19]. Four flasks were assigned to each of the 16 treatments. Cell toxicity is measured by the amount of lactic dehydrogenase (LHD) enzyme percentage leakage from each of the 64 flasks after 0.01, 0.25, 0.5, 1, 2, and 3 h since the application of the treatment. For illustration purposes, we consider the between-subjects factor CCl_4 levels 2.5 and 5.0 as fixed (new technology), and levels 0 and 1.0 are considered as existing random levels (old technology). Similarly, we consider the between-subjects factor $CHCl_3$ levels 10 and 25 as fixed (new technology), and levels 0 and 5 are taken as random levels (old technology). Of interest in the analysis is the percentage leakage observed at times 1, 2, and 3. We demonstrate the model construction procedure under certain assumptions in a completely randomised design (CRD) and in a repeated-measures design (RMD).

2.2. Construction of a Linear Mixed-Effects Model in CRD

Consider a three-way treatment structure in a balanced, completely randomised design (CRD) with full interaction of factors A , B , and C , each consisting of f fixed and r random levels. Assume we have f_A , f_B , and f_C fixed levels and r_A , r_B , and r_C random levels of factor A , respectively. We partition the dataset based on the combinations of factor levels and construct a partitioned model in each partition. For example, the FRF partitioned model is built from the f_A , r_B , and f_C levels. Similarly, other possible partitions are FFF, FFR, RFF, RRF, RFR, FRR, and RRR. We illustrate the model construction using the FRF linear

mixed-effects model in CRD, having at least one replication per treatment combination and expressed as

$$\mathbf{y}_{\text{FRF}_{ijkl}} = \mu_{\text{FRF}} + \varphi_{A_i} + \varphi_{B_j} + \varphi_{C_k} + \pi_1 + \dots + \pi_t + \epsilon_{\text{FRF}_{ijkl}}, \quad (1)$$

where $\mathbf{y}_{\text{FRF}_{ijkl}}$ is the l th observation in the (ijk) th treatment cell of the FRF partition; $l = 1, \dots, r_h$ are the replicates (where all $r_h = r$ for balanced data); μ_{FRF} is the overall mean; φ_{A_i} , φ_{B_j} , φ_{C_k} are the main effects of the three factors; π_1, \dots, π_t are the interaction effects; and φ_{A_i} ($i = 1, 2, \dots, f_A, f_A + 1, f_A + 2, \dots, a$ ($a = f_A + r_A$)), φ_{B_j} ($j = r_B + 1, r_B + 2, \dots, b$ ($b = f_B + r_B$)), and φ_{C_k} ($k = 1, 2, \dots, f_C, f_C + 1, f_C + 2, \dots, c$ ($c = f_C + r_C$)) are unknown parameters corresponding to fixed factor A , random factor B , and fixed factor C , respectively. Defining the random main effect as φ_R and the random interaction effect as π_R in (1), the random effects and the random error term ϵ_{ijkl} , are commonly assumed to have zero mean and variance, i.e., $\varphi_R \sim N(0, \sigma_{\varphi_R}^2)$, $\pi_R \sim N(0, \sigma_{\pi_R}^2)$, and $\epsilon_{ijkl} \sim N(0, \sigma_{\epsilon}^2)$.

For a balanced data scenario with r replications per cell, for example, the general linear mixed model Equation (1) is normally expressed in matrix form as

$$\mathbf{y}_{\text{FRF}} = \mathbf{X}_{\text{FRF}}\boldsymbol{\beta} + \mathbf{Z}_{\text{FRF}}\mathbf{u} + \boldsymbol{\epsilon}_{\text{FRF}}, \quad (2)$$

where \mathbf{y}_{FRF} : $N \times 1$ is a vector of response observations in the FRF partition; matrix \mathbf{X}_{FRF} : $N \times p$ is a known incidence matrix associated with the vector of p fixed-effects $\boldsymbol{\beta}$: $p \times 1$ in the model; matrix \mathbf{Z}_{FRF} : $N \times q$ is a known incidence matrix associated with the vector of q random-effects \mathbf{u} : $q \times 1$ in the model; and $\boldsymbol{\epsilon}_{\text{FRF}}$: $N \times 1$ is a vector of random errors. The usual assumption under this model is that the random effects are $\mathbf{u} \sim N(0, \mathbf{G})$, and the random residuals are $\boldsymbol{\epsilon}_{\text{FRF}} \sim N(0, \mathbf{R})$, where $\mathbf{R} = \sigma^2\mathbf{N}$ and \mathbf{G} is a diagonal matrix of variance components (i.e., different variances, and all zero covariances).

$$\mathbf{G} = \begin{bmatrix} \sigma_1^2 \mathbf{I}_r & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_t^2 \mathbf{I}_r \end{bmatrix} \text{ and } \mathbf{R} = \begin{bmatrix} \sigma_{\epsilon}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\epsilon}^2 \end{bmatrix}$$

The total variance–covariance is the structured matrix $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$, a structure that guarantees independence and homogeneity of residual errors. This implies that the variance of \mathbf{y} is modelled through \mathbf{Z} , \mathbf{G} , and \mathbf{R} . The simple total variance–covariance, \mathbf{V} , has a block-diagonal structure given by the matrix

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 \mathbf{I}_r + \sigma_{\epsilon}^2 \mathbf{I}_r & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_t^2 \mathbf{I}_r + \sigma_{\epsilon}^2 \mathbf{I}_r \end{bmatrix}$$

2.3. Linear Mixed-Effects Model in RMD

Traditionally, between-subject and within-subject factors in repeated-measures experiments are designated as either fully fixed effects or random effects. Depending on the objectives of the experiment, some factors in linear mixed-effects models may exist with both fixed and random levels [12,20,21]. The same scenario is common with a repeated-measures experiment, in which either the between-subjects factor or the repeated measures consist of both fixed and random levels. For instance, for improved results, a researcher may decide to consider additional levels of a between-subjects factor, in addition to the old and existing levels. In that case, the improved analysis needs to consider the new factor levels as fixed levels, and the old and existing levels are considered random. The approach creates an opportunity to compare and evaluate the effectiveness of new factor levels (fixed) against existing ones (random) and/or to compile a combined analysis of both. In addition, random levels allow for the assessment of variation between and within the factor levels in the entire population.

We consider a three-factor repeated-measures experiment with n experimental units (EU) that are randomly assigned to each of the a levels of the between-subjects factor A (with f_a fixed levels and r_a random levels, $a = f_a + r_a$, for example), the between-subjects factor B (with f_b fixed levels and r_b be random levels, $b = f_b + r_b$, for example), and the within-subjects factor C with t measurements (all considered to be fixed in this case) taken on each of the experimental units (EU).

$$y_{pijkl} = \mu_p + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(ij)} + \tau_l + (\alpha\tau)_{il} + (\beta\tau)_{jl} + (\alpha\beta\tau)_{ijl} + \epsilon_{pijkl}, \quad (3)$$

where the subscript p in y_{pijkl} and ϵ_{pijkl} denotes the partition ($p = 1, 2, \dots, 4$) and where y_{pijkl} is the l th measurement ($l = 1, 2, \dots, t$) of the k th experimental unit ($k = 1, 2, \dots, n$) in the i th level ($i = 1, 2, \dots, a$) of factor A and the j th level ($j = 1, 2, \dots, b$) of factor B in the p th partition. Depending on whether the effects are fixed or random, the model parameters are as defined in (1).

2.4. Model Assumptions

In experimental research that involves analysis of variance (ANOVA) as a technique for comparing different treatment means, a set of assumptions that includes the usual normality and homogeneity of variance must be checked before the analysis of data [22]. These traditional normality tests, such as Q–Q plots or the Shapiro–Wilk test, and outlier detection approaches (e.g., box plots), are appropriate for diagnosing violations of assumptions.

2.4.1. Sphericity (Circularity) Assumption

Similar to the homogeneity of variances in a between-subjects analysis of variance, sphericity holds in ANOVA for repeated-measures designs when the variances of the differences among all possible pairs of the within-subject factor group means are equal [23,24]. The assumption is usually unrealistic in repeated-measures designs in which observations are correlated. Univariate tests for within-subjects effects apply when sphericity holds. If the sphericity assumption is violated, several alternatives for adjusting the numerator and denominator degrees of freedom have been suggested [25]. The most popular degrees of freedom corrections are the Greenhouse–Geisser [26] or the Huynh–Feldt [27] adjustments, which are extensions of the Box correction factor [28]. These adjustments provide a more accurate adjusted p -value [29,30]. Failure to address the problem of sphericity when conducting an analysis of variance often leads to inflated F-ratios, type I errors, biased post hoc tests for group mean differences, and inaccurate conclusions [24]. Tests for sphericity are offered in most statistical computing packages (SAS, R, SPSS, etc.).

2.4.2. Compound Symmetry Assumption

An overly restrictive assumption closely related to sphericity is the compound symmetry assumption. It states that there is a constant variance and correlation between observations from the same subject, which is not always realistic in many repeated-measures applications [31]. The Greenhouse–Geisser [26] or the Huynh–Feldt [27] adjustments are used to circumvent restrictive compound symmetry assumptions and to accommodate more general covariance structures for the repeated measures [1]. Compound symmetry implies sphericity, but not vice versa. Univariate tests for within-subject effects apply when compound symmetry holds. If compound symmetry (and hence sphericity) holds, then a split-plot analysis is used as an appropriate approximation to the repeated-measures experiment, since it provides relatively more accurate p -values for testing treatment effects [31]. Compound symmetry is simplified in the Huynh–Feldt [27] condition for sphericity. Hence, both are tested in many software packages by Mauchly's test [32]. However, Mauchly's test for sphericity has been criticised for its over sensitivity and tendency to reject compound symmetry [33].

2.5. Estimation Techniques

The general linear mixed-effects model (2) is used to describe data from partitioned repeated measurements, wherein the fixed-effects component, $X_p\beta$, consists of the design matrix X_p , and the fixed-effects coefficients β are as defined in (3). The random-effects component, $Z_p\mathbf{u}$, contains the block-diagonal random-effects design matrix Z_p with design matrices for the individual subjects (Z_i); \mathbf{u} is a vector of random coefficients (the between-subject variance-covariance components), and ϵ_p denotes the within-subject errors from the p th partition. The random effects and the residuals follow the distribution $\mathbf{u} \sim N(0, \mathbf{G})$ and $\epsilon_p \sim N(0, \mathbf{R})$, respectively, where \mathbf{G} is a block-diagonal covariance matrix of the random effects and where \mathbf{R} is a diagonal covariance matrix with partitions corresponding to each subject (within-subject errors) in the analysis. The covariance matrix for the repeated-measures data is composed of matrices Z_p , \mathbf{G} , and \mathbf{R} , which is a block-diagonal $\Sigma_p = \text{Var}(\mathbf{y}_p) = Z_p\mathbf{G}Z_p' + \mathbf{R}$. The non-singular components \mathbf{G} and \mathbf{R} are usually estimated by two principal likelihood methods for estimating variance components [34], i.e., the maximum likelihood (ML) method and restricted maximum likelihood (REML). These procedures are available in various mixed-model statistical software, such as SAS (*PROC MIXED* procedure) and R, with the REML estimates generally preferred unless the data sets are quite large [34].

Assuming that both the random effects and the error terms are normally distributed, the likelihood function for the repeated-measures mixed model is given by [35]:

$$\begin{aligned} l &= \log [L(\mathbf{y}_p)] \\ &= \frac{-N}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_p| - \frac{1}{2} (\mathbf{y}_p - X_p\beta)' \Sigma_p^{-1} (\mathbf{y}_p - X_p\beta) \\ &= C - \frac{1}{2} \log |\Sigma_p| - \frac{1}{2} (\mathbf{y}_p - X_p\beta)' \Sigma_p^{-1} (\mathbf{y}_p - X_p\beta) \end{aligned} \quad (4)$$

where \mathbf{y}_p and $V = \Sigma_p$ are as defined in (2). Similarly, a modification of the ML procedure through factorisation of the likelihood function is proposed as an alternative method of estimating covariance parameters, which is the restricted maximum likelihood function [36]:

$$\begin{aligned} l_{re} &= \log [L(\mathbf{y}_p)] \\ &= \frac{-N}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_p| - \frac{1}{2} (\mathbf{y}_p - X_p\beta)' \Sigma_p^{-1} (\mathbf{y}_p - X_p\beta) \\ &= C - \frac{1}{2} \log |X_p' \Sigma_p^{-1} X_p| - \frac{1}{2} \log |\Sigma_p| - \frac{1}{2} (\mathbf{y}_p - X_p\hat{\beta})' \Sigma_p^{-1} (\mathbf{y}_p - X_p\hat{\beta}) \end{aligned} \quad (5)$$

where the available covariance matrix Σ_p is used to estimate the fixed-effects parameters,

$$\hat{\beta} = [X_p' \Sigma_p^{-1} X_p]^{-1} X_p' \Sigma_p^{-1} \mathbf{y}_p.$$

The main challenge in repeated-measures analysis of variance is to determine the adequate correlation structure, because the constant variance assumption for the distribution of the error terms is likely not to be reasonable for the distribution of error terms within subjects. There are various possible choices of covariance structures for repeated measures within each subject depending on the chosen parameterisation for \mathbf{G} and \mathbf{R} . The choices are usually guided by limitations of the software and by insight from the researcher. The most common covariance structures include variance components, compound symmetry (common covariance plus diagonal), unstructured (general covariance), and autoregressive [37]. With the *PROC MIXED* statement in SAS, one can specify any repeated-measurements covariance structure for \mathbf{G} by using the *Random* statement and by specifying the form of \mathbf{R} with the *Repeated* statement, in conjunction with the *Type* option [38]. Excluding the *Repeated* statement specifies the classical \mathbf{R} , which is assumed to be equal to $\sigma^2\mathbf{I}$.

There are numerous ways of identifying the most appropriate covariance structure amongst set candidate structures [37]. The most recommended approach is to select the structure that gives the smallest Akaike's Information Criterion (AIC) [39], a statistic that is defined by the model and the maximum likelihood estimates of the parameters from specifying the variance–covariance as

$$\text{AIC} = (-2)\mathcal{L}(\hat{\beta}, \hat{\Sigma}_p) + 2(g), \quad (6)$$

where g is the effective number of independently adjusted parameters in the covariance matrix from the p th partition, and $\mathcal{L}(\hat{\beta}, \hat{\Sigma}_p) = \log(ML)$ is the value of the likelihood function evaluated at $(\hat{\beta}, \hat{\Sigma}_p)$. A better model is the one with the smallest AIC value. Different forms of \mathbf{R} can be compared for adequacy using the likelihood ratio test statistic [38]. The hypotheses involved are the following: H_0 : \mathbf{R}_1 is as adequate as \mathbf{R}_2 ; and H_1 : \mathbf{R}_1 is not as adequate as \mathbf{R}_2 , where \mathbf{R}_1 is a special case of \mathbf{R}_2 . Suppose \mathbf{R}_1 and \mathbf{R}_2 have g_1 and g_2 parameters, respectively, with $(g_1 < g_2)$, and the test statistic is $Q = (-2)[\mathcal{L}(\hat{\beta}_1, \hat{\Sigma}_1) - \mathcal{L}(\hat{\beta}_2, \hat{\Sigma}_2)]$, which is distributed as $\chi^2(g_2 - g_1)$. We reject H_0 when $Q \geq \chi^2_{\frac{\alpha}{2}}(g_2 - g_1)$.

2.6. Methods of Inference

We present the algorithm for obtaining expected mean squares using the ANOVA approach for the FRF model.

2.6.1. Process for Deriving Expected Mean Squares

- (a) Based on the model involved, construct a two-way table with column headings corresponding to the source of variation, effect labels, each of the subscripts included in the model, and row headings corresponding to each source of variation in the ANOVA table.
- (b) Above each subscript, write the associated number of factor levels and insert on top either an "F" if the factor levels are fixed or an "R" if the factor levels are random.
- (c) Create an extra column on the extreme right for the variance components corresponding to the source of variation, and insert the appropriate random variance component (σ^2) or fixed variance component (θ) for each source of variation.
- (d) Compare the column subscript and the factor effect in each row, and write the number of levels corresponding to that subscript if the column subscript is not included in the factor effect label. Otherwise, leave it blank.
- (e) For rows that have effects that contain bracketed subscripts, write a "1" under the column if the subscript is included in the bracket.
- (f) For each row that has a fixed variance component (θ), put a zero in the cell headed by an "F" when the subscript is included in the effect label.
- (g) Enter a "1" in all remaining blank cells.
- (h) To obtain the expected mean squares for each effect, identify all the variance components associated with that effect label. Cover the column(s) headed by the effect subscript(s) in that effect, and obtain the coefficient of each of the identified components from the product of the entries in the column(s) headed by the uncovered subscript(s). Include the variance component σ_e^2 with the coefficient of 1 in the list.

Once the coefficients of variance components are established in step (m) of Figure 1, the expected mean squares are found by Step (h) of the Process in Section 2.6.1. Table 1 summarises the variance components obtained for a three-factor repeated-measures design, with two between-subjects factors (A and B) on one within-subject factor (C) using the Process in Section 2.6.1. We use this process to build an FRF model, which assumes factors A and C as fixed, factor B as random, and experimental units (EU) as random. The F, R, F, and Rep denote the effect; a , b , t , and n denote levels of factors A and B , treatment number, and sample size, respectively. The subscripts i , j , l , and k denote the indices of the corresponding levels.

Figure 1 illustrates the key steps to be taken when constructing the coefficients of variance components for a partitioned linear mixed-effects model.

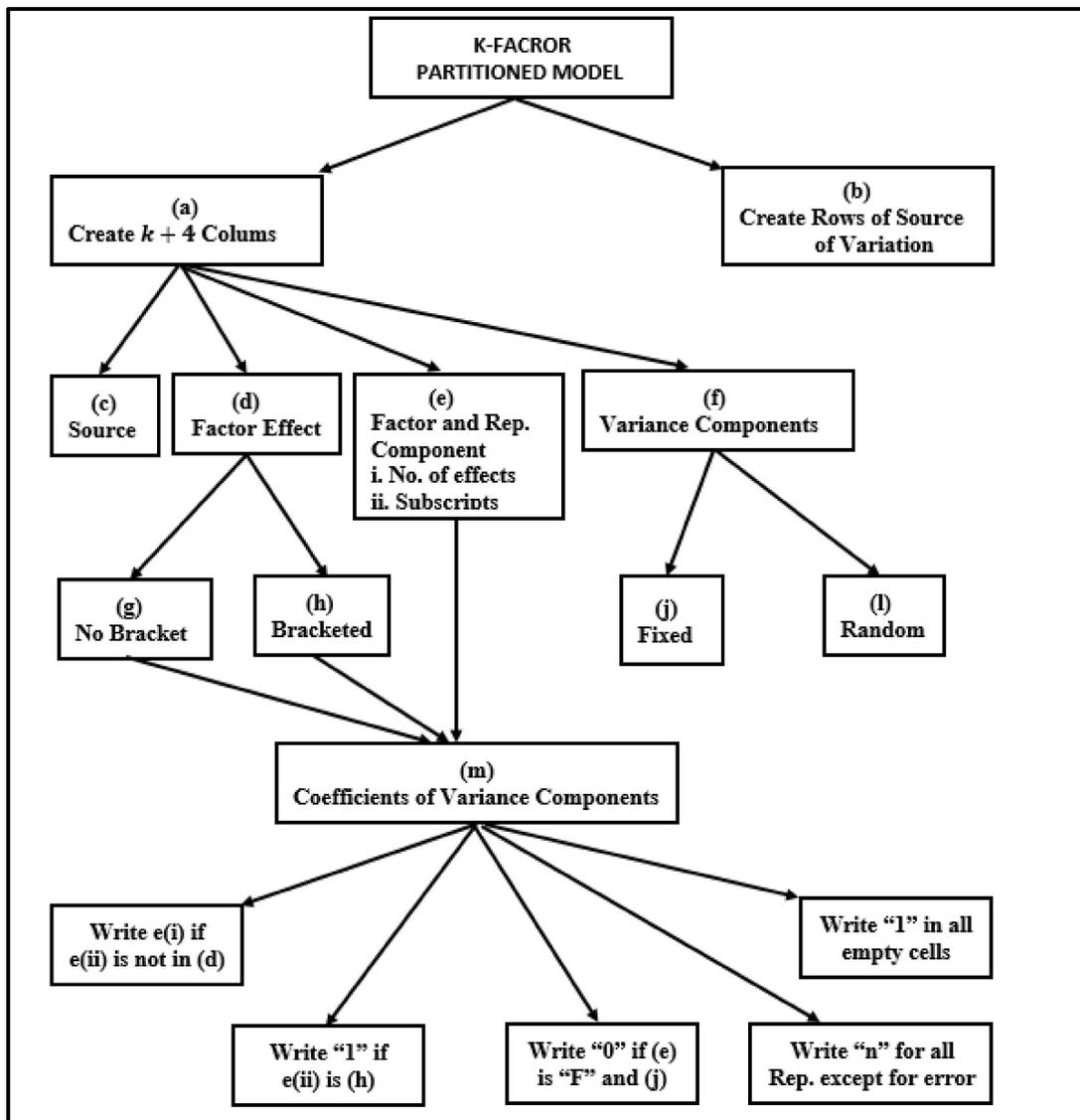


Figure 1. Process of constructing coefficients of variance components.

Table 1. Variance components for a three-factor repeated-measures model.

Source of Variation	Factor Effect	F	R	F	Rep	Components of Variance
		<i>a</i>	<i>b</i>	<i>t</i>	<i>n</i>	
		<i>i</i>	<i>j</i>	<i>l</i>	<i>k</i>	
<i>A</i>	α_i	0	<i>b</i>	<i>t</i>	N	θ_α
<i>B</i>	β_j	<i>a</i>	1	<i>t</i>	N	σ_β^2
<i>AB</i>	$\alpha\beta_{ij}$	1	1	<i>t</i>	N	$\sigma_{\alpha\beta}^2$
<i>EU</i>	$\gamma_{k(ij)}$	1	1	<i>t</i>	1	$\sigma_{\gamma(\alpha\beta)}^2$
<i>P</i>	τ_l	<i>a</i>	<i>b</i>	0	N	θ_τ
<i>PA</i>	$\alpha\tau_{il}$	0	<i>b</i>	0	N	$\theta_{\alpha\tau}$
<i>PB</i>	$\beta\tau_{jl}$	<i>a</i>	1	1	N	$\sigma_{\beta\tau}^2$
<i>PAB</i>	$\alpha\beta\tau_{ijl}$	1	1	1	N	$\sigma_{\alpha\beta\tau}^2$
Error	$\epsilon_{l(ijk)}$	1	1	1	1	σ_ϵ^2

For example, $E(MSA)$, with effect α_i , is composed of the variance components $\theta_\alpha, \sigma_{\alpha\beta}^2, \sigma_{\gamma(\alpha\beta)}^2, \theta_{\alpha\tau}, \sigma_{\alpha\beta\tau}^2$, and σ_ϵ^2 as follows:

$$E(MSA) = \sigma_\epsilon^2 + t\sigma_{\gamma(\alpha\beta)}^2 + tn\sigma_{\alpha\beta}^2 + n\sigma_{\alpha\beta\tau}^2 + btn\theta_\alpha.$$

Table 2 displays the ANOVA layout and the expected mean squares for a three-factor repeated-measures design when one of the factors is a within-subject factor. The FRF model is considered for illustration, with experimental units assumed to be random.

Table 2. Expected mean squares for a three-factor repeated-measures design.

Source of Variation	Sum of Squares	Degrees of Freedom	E(MS)
A	SSA	$a - 1$	$\sigma_\epsilon^2 + t\sigma_{\gamma(\alpha\beta)}^2 + tn\sigma_{\alpha\beta}^2 + n\sigma_{\alpha\beta\tau}^2 + btn\theta_\alpha$
B	SSB	$b - 1$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta\tau}^2 + an\sigma_{\beta\tau}^2 + t\sigma_{\gamma(\alpha\beta)}^2 + tn\sigma_{\alpha\beta}^2 + ant\sigma_\beta^2$
$A \times B$	SS($A \times B$)	$(a - 1)(b - 1)$	$\sigma_\epsilon^2 + t\sigma_{\gamma(\alpha\beta)}^2 + n\sigma_{\alpha\beta\tau}^2 + tn\sigma_{\alpha\beta}^2$
Unit($A \times B$)	SSU($A \times B$)	$ab(n - 1)$	$\sigma_\epsilon^2 + t\sigma_{\gamma(\alpha\beta)}^2$
Period	SSP	$t - 1$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta\tau}^2 + an\sigma_{\beta\tau}^2 + abn\theta_\tau$
Period \times A	SSP \times A	$(a - 1)(t - 1)$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta\tau}^2 + bn\theta_{\alpha\tau}$
Period \times B	SSP \times A	$(b - 1)(t - 1)$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta\tau}^2 + an\sigma_{\beta\tau}^2$
Period \times A \times B	SSP \times A \times B	$(a - 1)(b - 1)(t - 1)$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta\tau}^2$
Residual	SSE	$ab(t - 1)(n - 1)$	σ_ϵ^2
Total	SST	$abtn - 1$	

Note that: $\theta_\alpha = \frac{1}{a-1} \sum_{i=1}^a \alpha_i^2$; $\theta_{\alpha\tau} = \frac{1}{(a-1)(t-1)} \sum_{i=1}^a \sum_{j=1}^t (\alpha\tau)_{ij}^2$, and $\theta_\tau = \frac{1}{t-1} \sum_{i=1}^t \tau_i^2$.

2.6.2. Hypothesis Testing for Fixed Effects

We are interested in testing the main and interaction effects of the between-subjects factor and of the within-subjects factor in both the partitioned and combined repeated-measures linear mixed-effects model. In addition to checking model assumptions, the following hypotheses are of interest for each partitioned model:

Hypothesis 1 (H1). *Between-subjects main and interaction effects (i.e., $H_0 : \alpha_i = 0$).*

Hypothesis 2 (H2). *Within-subjects main and interaction effects (i.e., $H_0 : \tau_i = 0$).*

The test statistic for H1 is given by $F = \frac{MSA}{MSU(A \times B)} \sim F_{a-1, ab(n-1)}(\alpha)$, and the test statistic for H2 is given by $F = \frac{MSP}{MSE} \sim F_{t-1, ab(t-1)(n-1)}(\alpha)$. The interaction effects among the between-subjects and the interaction effects among within-subjects are tested by the $MSU(A \times B)$ and the MSE on the denominator, respectively.

2.6.3. Hypothesis Testing for Random Effects

Variance components are estimated by equating the mean square to the expected mean squares derived in Table 2. Where there are no valid F tests, approximate F tests are constructed for the sources of variability in random effects [40]. For random factor B, the hypothesis of interest may be

Hypothesis 3 (H3). *Random effects (e.g., $H_0 : \sigma_\beta^2 = 0$, against $H_1 : \sigma_\beta^2 > 0$).*

2.6.4. Combined Analysis

The individual partitioned models provide pieces of information which are needed for an integrated analysis. The combined model is built by combining the degrees of

freedom and sum of squares associated with each source of variation for each appropriate hypothesis test. For example, the combined effect of the within-subjects factor (*Period*) in the FC model is obtained from the partitions in which the factor *Period* is fixed, i.e., the pieces of information are supplied by the partitioned models FFF, FRF, RFF, and RRF. Similarly, the other main and interaction effects for the combined model are obtained by summing up the associated degrees of freedom and sums of squares.

3. Results

3.1. Checking Model Assumptions

A three-factor (CCl_4 , $CHCl_3$, and *Time*) repeated-measures experiment on one of the factors (*Time*) is proposed. Table 3 shows the multivariate data (wide format) layout for a three-factor repeated-measures experiment.

Table 3. Data layout for a three-factor repeated-measures experiment.

Factors			Time Period				
CCl_4	$CHCl_3$	Flask	1	2	...	t	
1	1	1	y_{1111}	y_{1112}	...	y_{111t}	
		⋮	⋮	⋮	⋮	⋮	
		n	y_{11n1}	y_{11n2}	...	y_{11nt}	
	2	1	y_{1211}	y_{1212}	...	y_{121t}	
		⋮	⋮	⋮	⋮	⋮	
		n	y_{12n1}	y_{12n2}	...	y_{12nt}	
	⋮	⋮	⋮	⋮	⋮	⋮	
	b	1	y_{1b11}	y_{1b12}	...	y_{1b1t}	
		⋮	⋮	⋮	⋮	⋮	
		n	y_{1bn1}	y_{1bn2}	...	y_{1bnt}	
	2	1	1	y_{2111}	y_{2112}	...	y_{211t}
			⋮	⋮	⋮	⋮	⋮
n			y_{21n1}	y_{21n2}	...	y_{21nt}	
2		1	y_{2211}	y_{2212}	...	y_{221t}	
		⋮	⋮	⋮	⋮	⋮	
		n	y_{22n1}	y_{22n2}	...	y_{22nt}	
⋮		⋮	⋮	⋮	⋮	⋮	
b		1	y_{2b11}	y_{2b12}	...	y_{2b1t}	
		⋮	⋮	⋮	⋮	⋮	
		n	y_{2bn1}	y_{2bn2}	...	y_{2bnt}	
⋮		⋮	⋮	⋮	⋮	⋮	
a		1	1	y_{a111}	y_{a112}	...	y_{a11t}
	⋮		⋮	⋮	⋮	⋮	
	n		y_{a1n1}	y_{a1n2}	...	y_{a1nt}	
	2	1	y_{a211}	y_{a212}	...	y_{a21t}	
		⋮	⋮	⋮	⋮	⋮	
		n	y_{a2n1}	y_{a2n2}	...	y_{a2nt}	
	⋮	⋮	⋮	⋮	⋮	⋮	
	b	1	y_{ab11}	y_{ab12}	...	y_{ab1t}	
		⋮	⋮	⋮	⋮	⋮	

3.1.1. Normality and Outliers

Q–Q plots are used to check normality and outlier assumptions simultaneously. The FFF, FRF, RFF, and RRF data sets do not show any serious deviations from normality. Figure 2 shows the normal Q–Q plots for the four data subsets.

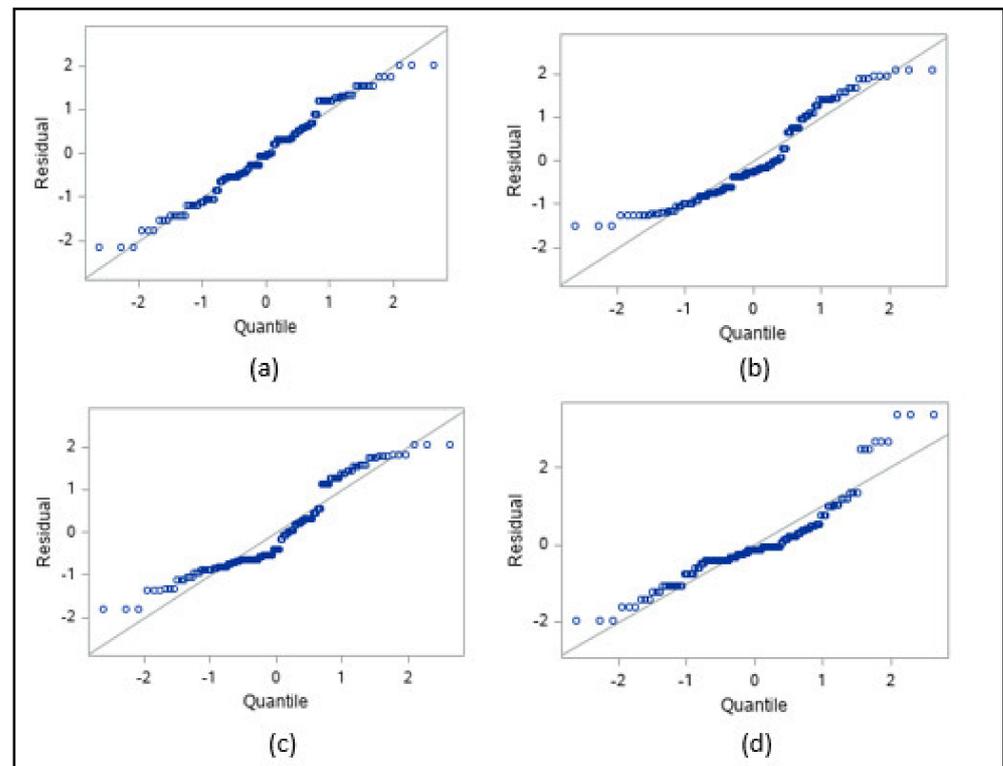


Figure 2. Q–Q plots for (a) FFF, (b) FRF, (c) RFF, and (d) RRF data sets.

Furthermore, the Q–Q plots do not show any influential point (outlier) that warrants exclusion, since all plots are not very far from the diagonal.

3.1.2. Sphericity and/or Compound Symmetry

In order to test the suitability of using a repeated-measures design in the experiment, the sphericity (or compound symmetry) assumption is tested for each partitioned data subset. Sphericity test results are produced by fitting two models using the *PROC MIXED* procedure: one specifying the unrestricted covariance structure and the other with a less conservative Huynh–Feldt (H–F) adjustment in the *Type* option [41]. The null hypothesis for the test is:

Hypothesis 4 (H4). *The covariance structure fits the sphericity structure.*

The difference (D) between the two $-2 \log$ -likelihoods of the two competing models follows a Chi-square distribution with degrees of freedom equal to the difference in the numbers of parameters in them.

FFF model: $\chi_{35}^2(0.05) \approx 55.76$, $df = 35$, $D = 1722.6$, significant;

FRF: $\chi_{38}^2(0.05) \approx 55.76$, $df = 38$, $D = 184.54$, significant;

RFF: $\chi_{37}^2(0.05) \approx 55.76$, $df = 37$, $D = 288.8$, significant;

RRF: $\chi_{36}^2(0.05) \approx 55.76$, $df = 35$, $D = 384.1$, significant.

The sphericity assumption H4 fails in all the four (4) partitioned data sets.

3.2. Analysis of Results

We use the restricted maximum likelihood estimation (REML) approach to estimate these variance components. Table 4 contains the estimated AIC from each of the covariance structures as well as the number of covariance parameters estimated with a non-zero value.

Table 4. Akaike’s Information Criteria (AIC) for the covariance structures in partitioned models.

Model FFF			Model FRF		
Covariance Structure	Number of Parameters	AIC	Covariance Structure	Number of Parameters	AIC
CS	2	−282.6	CS	4	−361.3
AR(1)	2	−443.1	AR(1)	3	−492.4
ARH(1)	10	−525.6	ARH(1)	13	−506.3
CSH	10	−356.7			
Model RFF			Model RRF		
Covariance Structure	Number of Parameters	AIC	Covariance Structure	Number of Parameters	AIC
CS	4	−450.1	CS	4	−665.8
AR(1)	4	−522.4	AR(1)	4	−782.4
ARH(1)	13	−509.3	ARH(1)	13	−785.3

AR(1): First-order autoregressive; ARH(1): Heterogeneous first-order autoregressive; CS: Compound symmetry; VC: Variance components; CSH: Heterogeneous compound symmetry.

The covariance structure ARH(1) has the minimum AIC values in models FFF, FRF, and RRF, and covariance structure AR(1) has the smallest AIC value in model RFF. The covariance structure ARH(1) is chosen as the most adequate covariance structure for the partitioned models based on AIC.

When the appropriate covariance structure is incorporated into the model, the relationships among the errors and variance components are specified. The PROC MIXED procedure does not require further assumptions on them [42], making it a robust and flexible procedure [15] that produces valid F-tests regardless of whether the sphericity assumption is satisfied or not. Table 5 summarises the PROC MIXED F-tests for the main and interaction effects of the between-subjects and the within-subjects factors for the four partitioned models.

Table 5. Type III tests of fixed effects of FFF, FRF, FFR, and FRR models.

Model	Effect	Numerator Degrees of Freedom	Denominator Degrees of Freedom	F	p-Value
FFF	CCl_4	1	12	0.24	0.6363
	$CHCl_3$	1	12	17.33	0.0013 **
	$CCl_4 \times CHCl_3$	2	12	1.35	0.2678
	Time	2	24	95.99	<0.0001 **
	Time \times CCl_4	2	24	2.66	0.0908
	Time \times $CHCl_3$	2	24	20.91	<0.0001 **
	Time \times $CCl_4 \times CHCl_3$	2	24	1.26	0.3023
FRF	$CHCl_3$	1	1	0.5	0.6079
	Time	2	2	2.17	0.3156
	Time \times CCl_4	2	2	2.83	0.2610
RFF	$CHCl_3$	1	1	7.07	0.2290
	Time	2	2	5.42	0.1557
	Time \times $CHCl_3$	2	2	3.26	0.2348
RRF	Time	2	2	1.01	0.4978

*** significant at $\alpha = 0.05$.

The $CHCl_3$ factor has a significant effect on the leakage percentage (p -value < 0.001) over time in the FFF partition. Based on the fixed-effects (FFF) repeated-measures study, it may be concluded that chloroform ($CHCl_3$) has a significant impact on the amount of lactic dehydrogenase (LHD) enzyme percentage leakage (toxicity of cells) over time, whereas neither carbon tetrachloride (CCl_4) in isolation nor the interaction thereof has significant influence. Furthermore, the *Time* factor plays an important role in determining the amount of leakage as well. However, the interaction of old (random) and new (fixed) levels of the between-subjects factor levels has non-significant effects in FRF, RFF, and RRF models when $CHCl_3$ fixed levels are involved.

Table 6 gives a summary of the estimated covariance parameters and the proportion (in percentage) of variation they contribute in each of the partitioned models. A zero variance for the random levels of chloroform ($CHCl_3$) and very small estimates in other factors are obtained, which result in very low estimates of covariance parameters in the models.

Table 6. Covariance parameter estimates of FRF, RFF, and RRF, models based on ARH(1) covariance structure.

Model	Covariance Parameter	Estimate	Standard Error	Proportion of Variation Accounted for
FRF	$CHCl_3$	0	0	0
	$CCl_4 \times CHCl_3$	0.04249	0.04470	22.5
	$Time \times CHCl_3$	0	0	0
	$Time \times CCl_4 \times CHCl_3$	0.000621	0.000759	0.3
RFF	CCl_4	0.000607	0.003056	0.05
	$CCl_4 \times CHCl_3$	0	0	0
	$Time \times CCl_4$	0.000149	0.000809	0.02
	$Time \times CCl_4 \times CHCl_3$	0.000806	0.000984	0.08
RRF	CCl_4	0.0011	0.00401	2.6
	$CHCl_3$	0.00306	0.00677	7.1
	$CCl_4 \times CHCl_3$	0.00232	0.00396	5.4
	$Time \times CCl_4$	0.000002	0.00002	0.0
	$Time \times CHCl_3$	0.000065	0.000096	0.1
	$Time \times CCl_4 \times CHCl_3$	0.000013	0.000025	0.0

The $CHCl_3$ and CCl_4 random levels and their interaction have a noticeable contribution to the proportion of variation in the amount of lactic dehydrogenase (LHD) enzyme percentage leakage in the RRF partition. Generally, *Time* has very little interaction effect with the between-subjects factors in FRF and RRF models in determining the proportion of variation of the toxicity of cells.

Since the mixed-model methodology directly computes neither sums of squares nor F-statistics from the ratio of mean squares, and since the R CRAN package *lme4* for mixed models currently does not have options for other covariance structures to cater for correlated error variances, generating a combined analysis is not a straightforward exercise. We scrap the targeted data subset based on the effects of interest before fitting the models. Analogous to the partitioned analyses, a comparison of model fit via the AIC approach is conducted. For convenience purposes, let the factors CCl_4 , $CHCl_3$, and *Time* be labelled as factor *A*, *B*, and *C*, respectively. The PROC MIXED procedure is used to fit the repeated-measures linear mixed models for intended narrow inference spaces [12,43]. Table 7 shows the Type III tests for the combined models.

Of the possible candidate covariance structures (CS, CSH, AR(1), and ARH(1)), structure ARH(1) is selected as the most appropriate covariance structure for the combined fixed-effects model FA, and AR(1) is appropriate for FB and FA × FB. The factors *B* ($CHCl_3$) and *C* (*Time*) have significant effects (p -value < 0.05) in the combined models. The broad inference scope results for the combined models (assuming random factor *A* or *B* effects) are similarly analysed.

Table 7. Fixed-effects F-tests for the combined models in narrow inference scope.

Type III Tests of Fixed Effects in Combined Models						
Model	AIC [CS]	Effect	Num DF	Den DF	F	Pr > F
FA	−127.6 [ARH(1)]	A	1	24	3.36	0.0794
		B	3	24	5.81	0.0039 **
		A × B	3	24	9.10	0.0003 **
		C	2	48	12.33	0.0001 **
		A × C	2	48	2.52	0.0908
		B × C	6	48	2.33	0.0468 **
		A × B × C	6	48	0.76	0.6027
FB	−127.6 [AR(1)]	A	3	24	0.53	0.6631
		B	1	24	16.3	0.0005 **
		A × B	3	24	0.22	0.8801
		C	2	48	24.82	<0.0001 **
		A × C	6	48	1.09	0.3795
		B × C	2	48	7.82	0.0012 **
		A × B × C	6	48	0.50	0.8064
FA × FB	−69.1 [AR(1)]	A	1	12	0.14	0.7106
		B	1	12	10.63	0.0068 **
		A × B	1	12	0.83	0.3807
		C	2	24	19.75	<0.0001 **
		A × C	2	24	0.58	0.5667
		B × C	2	24	4.65	0.0196 **
		A × B × C	2	24	0.26	0.7725

*** significant at $\alpha = 0.05$.

4. Discussion

Based on the illustrative example results, the approach manages to isolate the effects of new and old factor levels over time. The combined analysis confirmed the results of the partitioned analysis on the percentage leakage in cells. The proposed approach conforms to the model construction and the analysis procedures in a repeated-measures design. It can be used as a planning tool in which factor combination and time are of interest in designing experiments that involve repeated measures. In such experiments, blindly adopting the assumption of homogeneous error terms without exploring possible candidate covariance structures may compromise the ability of an experiment to detect sufficient variation in the response variable. In addition, our approach enhances the accuracy of inferences by providing partitioned analysis of heterogeneous variances and covariance structures, which sometimes are not identical in the data subsets.

Given the increased complexity of research data in various research fields, the application of a linear mixed-model methodology must be in line with data covariance structures for accurate results to be achieved. One of the approaches that has proved to be a reliable tool for managing big data complexity issues is the partitioning approach [12,20,21], in which the traditional homogeneous error variance structure is assumed. The current study extends the new approach to a three-factor treatment structure in a repeated-measures design in which linear mixed-effects models are applicable. In essence, the approach can be extended to cater for repeated-measures experiments in which any number of between-subjects and within-subjects factors are involved. In most cases, repeated-measures ex-

periments do not assume equal and uncorrelated error vectors, since regularly timed measurements taken on the same subject over time are usually correlated [34].

For the fixed-effects partitions, the linear mixed-effects models for a repeated-measures design are fit by the *PROC GLM* procedure, and a combined analysis of these can be obtained by syncretising the sum of squares and degrees of freedom from the fit models. However, obtaining a combined analysis using the SAS *PROC MIXED* procedure is impossible using the sum of squares approach, since the *PROC MIXED* procedure uses a likelihood-based estimation scheme instead of the least-squares method. A comparable alternative to reduction in the sum of squares for the fit model in *PROC MIXED* is to consider the amount of information retained by the fit model when compared to the null model.

5. Conclusions

The proposed approach allows for construction and hypothesis testing in repeated-measures data if a heterogeneous error structure is assumed. Although the MLE method used in SAS *PROC MIXED* does not estimate sums of squares, data scrapping based on targeted factor levels work equally well as an alternative approach to obtain the combined analysis. The proposed approach can be adopted as an essential tool for the comparison of new inventions against existing strategies and equipment. It leads to an exploration of the fixed and random levels of the same factor and the subsequent interaction of levels of factors of interest. We can assess the differences between levels of the same factor and understand variation within the same factor. In addition, modelling allows for the assessment of various covariance structures. Although this paper focuses on a limited scope of covariance structures, we pose an open research problem on the application of the proposed approach to other designs of experiments that incorporate more complex and non-diagonal covariance structures.

Author Contributions: Conceptualisation, L.C. and P.N.; methodology, L.C.; software, L.C.; validation, L.C. and P.N.; formal analysis, L.C.; writing—original draft preparation, L.C.; writing—review and editing, P.N.; supervision, P.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received publication funding from Sol Plaatje University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data for this research are available in [19,31].

Acknowledgments: The authors would like to acknowledge the support from the University of South Africa, and from the Sol Plaatje University for the publication costs provided.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AIC	Akaike's Information Criteria
ANOVA	Analysis of variance
AR(1)	First-order autoregressive
ARH(1)	Heterogeneous first-order autoregressive
CCl_4	Carbon tetrachloride
$CHCl_3$	Chloroform
CRD	Completely randomised design
CS	Compound symmetry
CSH	Heterogeneous compound symmetry
EU	Experimental unit
FFF	Fixed-Fixed-Fixed
FFR	Fixed-Fixed-Random
FRF	Fixed-Random-Fixed

FRR	Fixed-Random-Random
LHD	Lactic dehydrogenase
REML	Restricted maximum likelihood
RFF	Random-Fixed-Fixed
RFR	Random-Fixed-Random
RRF	Random-Random-Fixed
RRR	Random-Random-Random
SAS	Statistical Analysis Systems
VC	Variance components

Appendix A

SAS Code for Mixed-Model Analysis

```

/* Fitting FFR Model in Proc Mixed */
FILENAME REFFILE '/home/u35581214/LDH Leakage Data.sav';
PROC IMPORT DATAFILE = REFFILE
      DBMS = SAV
      OUT = LDH;
RUN;

/* View Repeated-Measures Data in Multivariate Form */
Proc print data = LDH;
run;

/* Set Repeated-Measures Data to Univariate form */
Data LDH_mult(keep = CCl4 CHCl3 Flask Time4 Time5 Time6)
LDH_univ(keep = CCl4 CHCl3 Flask Time Leakage);
set LDH;
output LDH_mult;
Leakage = Time4;Time = 1; output LDH_univ;
Leakage = Time5;Time = 2; output LDH_univ;
Leakage = Time6;Time = 3; output LDH_univ;
run;

/* View Data in Univariate and Multivariate Form */
Proc print data = LDH_univ;
run;
Proc print data = LDH_mult;
run;

/* Subset or partition FRF from LHD univariate original Data */
Data FRF;
set LDH_univ;
if (CCl4 = 2.5 AND CHCl3 = 0) then output;
if (CCl4 = 2.5 AND CHCl3 = 0) then output;
if (CCl4 = 2.5 AND CHCl3 = 0) then output;
if (CCl4 = 2.5 AND CHCl3 = 5) then output;
if (CCl4 = 2.5 AND CHCl3 = 5) then output;
if (CCl4 = 2.5 AND CHCl3 = 5) then output;
if (CCl4 = 5 AND CHCl3 = 0) then output;
if (CCl4 = 5 AND CHCl3 = 0) then output;
if (CCl4 = 5 AND CHCl3 = 0) then output;
if (CCl4 = 5 AND CHCl3 = 5) then output;
if (CCl4 = 5 AND CHCl3 = 5) then output;
if (CCl4 = 5 AND CHCl3 = 5) then output;
run;
Proc print data = FRF;

```

```

run;

/* Plot differences in leakages contributed by predictors */
proc means noprint data = FRF nway;
var Leakage;
class CCl4 CHCl3 Flask Time;
output out = avgFRF mean = avgLeakage;
run;
proc print data = avgFRF;
run;

/* New data set called avg created*/
/* Plot differences in leakage by predictor CHCl3 and Time */
Proc gplot data = avgFRF;
plot avgLeakage*Time = CHCl3/haxis = 0 to 8 by 1 hminor = 0 vminor = 0;
symbol1 v = star c = blue i = join l = 1;
symbol2 v = plus c = red i = join l = 2;
title "Percentage leakage per time per CHCl3";
run; Quit;

/* Partitioning FRF multivariate data for covariance analysis */
Data FRF_mult;
set LDH_mult;
if (CCl4 = 2.5 AND CHCl3 = 0) then output;
if (CCl4 = 2.5 AND CHCl3 = 0) then output;
if (CCl4 = 2.5 AND CHCl3 = 0) then output;
if (CCl4 = 2.5 AND CHCl3 = 5) then output;
if (CCl4 = 2.5 AND CHCl3 = 5) then output;
if (CCl4 = 2.5 AND CHCl3 = 5) then output;
if (CCl4 = 2.5 AND CHCl3 = 5) then output;
if (CCl4 = 2.5 AND CHCl3 = 5) then output;
if (CCl4 = 5 AND CHCl3 = 0) then output;
if (CCl4 = 5 AND CHCl3 = 0) then output;
if (CCl4 = 5 AND CHCl3 = 0) then output;
if (CCl4 = 5 AND CHCl3 = 5) then output;
if (CCl4 = 5 AND CHCl3 = 5) then output;
if (CCl4 = 5 AND CHCl3 = 5) then output;
run;
Proc print data = FRF_mult;
run;

/* Sphericity Test using PROC MIXED */
/* Sphericity test H0: Sphericity holds */
proc mixed data = FRF method = reml cl ic covtest;
class CCl4 CHCl3 Time Flask;
model Leakage = CCl4 | CHCl3;
random CHCl3 CCl4*CHCl3 CHCl3*Time CCl4*CHCl3*Time /s;
repeated/ subject = Flask(CCl4*CHCl3) type = un;
run;

proc mixed data = FRF method = reml cl ic covtest;
class CCl4 CHCl3 Time Flask;
model Leakage = CCl4 | CHCl3;
random CHCl3 CCl4*CHCl3 CHCl3*Time CCl4*CHCl3*Time /s;
repeated/ subject = Flask(CCl4*CHCl3) type = HF;
run;

/* Normality Q-Q plots */
ods graphics on;
proc mixed data = FRF plots = influenceestplot;

```

```

class CCl4 CHCl3 Time Flask;
model Leakage = CCl4 Time CCl4*Time/residual;
random CHCl3 CCl4*CHCl3 CHCl3*Time CCl4*CHCl3*Time;
repeated/subject = Flask(CCl4*CHCl3) type = cs r;
run;
ods graphics off;

/* Checking Covariance Structure */
proc corr data = FRF_mult cov;
var Time4 Time5 Time6;
run;

/* Fit the model by PROC MIXED and compare covariance structures */
proc mixed data = FRF method = reml cl ic covtest;
class CCl4 CHCl3 Time Flask;
model Leakage = CCl4 Time CCl4*Time/s;
random CHCl3 CCl4*CHCl3 CHCl3*Time CCl4*CHCl3*Time;
repeated/subject = Flask(CCl4*CHCl3) type = cs r;
run;

proc mixed data = FRF method = reml cl ic covtest;
class CCl4 CHCl3 Time Flask;
model Leakage = CCl4 Time CCl4*Time;
random CHCl3 CCl4*CHCl3 CHCl3*Time CCl4*CHCl3*Time/s;
repeated/subject = Flask(CCl4*CHCl3) type = arh(1) r;
lsmeans CCl4/pdiff cl adjust = tukey;
run;

proc mixed data = FRF method = reml cl ic covtest;
class CCl4 CHCl3 Time Flask;
model Leakage = CCl4 Time CCl4*Time;
random CHCl3 CCl4*CHCl3 CHCl3*Time CCl4*CHCl3*Time/s;
repeated/subject = Flask(CCl4*CHCl3) type = ar(1) r;
lsmeans CCl4/pdiff cl adjust = tukey;
run;

/* Scrapping Data for the Combined Model FB */
Data FB_univ;
set LDH_univ;
if (CHCl3 = 10) then output;
if (CHCl3 = 25) then output;
run;

/* Scrapping Data for the FA x FB Combined Model */
Data FAFB_univ;
set LDH_univ;
if (CCl4 = 2.5 and CHCl3 = 10) then output;
if (CCl4 = 2.5 and CHCl3 = 25) then output;
if (CCl4 = 5 and CHCl3 = 10) then output;
if (CCl4 = 5 and CHCl3 = 25) then output;
run;
Proc print data = FB_univ;
run;

/* Fitting the combined model FB (for narrow inferential scope) */
proc mixed data = FB_univ method = reml cl ic covtest;
class CCl4 CHCl3 Time Flask;
model Leakage = CCl4 CHCl3 CCl4*CHCl3 Time CCl4*Time CHCl3*Time CCl4*CHCl3*Time/s;

```

```

repeated/subject = Flask(CCI4*CHCI3) type = arh(1) r;
run;

/* Fitting the combined model FB (for broad inferential scope) */
proc mixed data = FB_univ method = reml cl ic covtest;
class CCI4 CHCI3 Time Flask;
model Leakage = CHCI3 Time CHCI3*Time/s;
random CCI4 CCI4*CHCI3 CCI4*Time CCI4*CHCI3*Time;
repeated/subject = Flask(CCI4*CHCI3) type = ar(1) r;
run;

```

References

1. Fitzmaurice, G.M.; Davidian, M.; Verbeke, G.; Molenberghs, G. (Eds.) *Longitudinal Data Analysis*; Chapman and Hall/CRC Handbooks of Modern Statistical Methods; Chapman & Hall/CRC: Boca Raton, FL, USA, 2009.
2. Pan, J.; Shang, J.A. Simultaneous Variable Selection Methodology for Linear Mixed Models. *J. Stat. Comput. Simul.* **2018**, *88*, 3323–3337. [[CrossRef](#)]
3. Demidenko, E. *Mixed Models: Theory and Applications with R*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 2013.
4. Verbeke, G.; Molenberghs, G. *Linear Mixed Models for Longitudinal Data*; Springer Series in Statistics; Springer: New York, NJ, USA; Berlin, Heidelberg, 2000.
5. Verbeke, G. The Linear Mixed Model. A Critical Investigation in the Context of Longitudinal Data Analysis. Ph.D. Dissertation, Catholic University of Leuven, Faculty of Sciences, Department of Mathematics, Leuven, Belgium, 1995.
6. Verbeke, G.; Lesaffre, E. The Effect of Misspecifying the Random-Effects Distribution in Linear Mixed Models for Longitudinal Data. *Comput. Stat. Data Anal.* **1997**, *23*, 541–556. [[CrossRef](#)]
7. Davis, C.S. *Statistical Methods for the Analysis of Repeated Measurements*; Springer Texts in Statistics; Springer: New York, NY, USA, 2002.
8. Muller, K.E.; Edwards, L.J.; Simpson, S.L.; Taylor, D.J. Statistical Tests with Accurate Size and Power for Balanced Linear Mixed Models. *Statist. Med.* **2007**, *26*, 3639–3660. [[CrossRef](#)] [[PubMed](#)]
9. Hickey, G.L.; Mokhles, M.M.; Chambers, D.J.; Kolamunnage-Dona, R. Statistical Primer: Performing Repeated-Measures Analysis. *Interact. Cardiovasc. Thorac. Surg.* **2018**, *26*, 539–544. [[CrossRef](#)] [[PubMed](#)]
10. Matuschek, H.; Kliegl, R.; Vasishth, S.; Baayen, H.; Bates, D. Balancing Type I Error and Power in Linear Mixed Models. *J. Mem. Lang.* **2017**, *94*, 305–315. [[CrossRef](#)]
11. Fitzmaurice, G.M.; Laird, N.M.; Ware, J.H. *Applied Longitudinal Analysis*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 2011.
12. Chaka, L.; Njuho, P. Construction of a Linear Mixed Model with Each Factor Having Both Fixed and Random Levels: A Case of Split-Split-Plot Structure in a RCBD. *Int. J. Agric. Stat. Sci.* **2021**, *17*, 501–518. Available online: <https://connectjournals.com/03899.2021.17.501> (accessed on 20 January 2022).
13. Crowder, M.J.; Hand, D.J. *Analysis of Repeated Measures*, 1st ed.; Monographs on Statistics and Applied Probability; Chapman and Hall: London, UK; New York, NY, USA, 1990.
14. Barnett, A.G.; Koper, N.; Dobson, A.J.; Schmiegelow, F.; Manseau, M. Using Information Criteria to Select the Correct Variance-Covariance Structure for Longitudinal Data in Ecology: Selecting the Correct Variance-Covariance. *Methods Ecol. Evol.* **2010**, *1*, 15–24. [[CrossRef](#)]
15. Wang, Z.; Goonewardene, L.A. The Use of MIXED Models in the Analysis of Animal Experiments with Repeated Measures Data. *Can. J. Anim. Sci.* **2004**, *84*, 1–11. [[CrossRef](#)]
16. Patterson, H.D.; Thompson, R. Recovery of Inter-Block Information When Block Sizes Are Unequal. *Biometrika* **1971**, *58*, 545–554. [[CrossRef](#)]
17. Diffey, S.M.; Smith, A.B.; Welsh, A.H.; Cullis, B.R. A New REML (Parameter Expanded) EM Algorithm for Linear Mixed Models. *Aust. N. Z. J. Stat.* **2017**, *59*, 433–448. [[CrossRef](#)]
18. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Society. Ser. B (Methodol.)* **1977**, *39*, 1–22.
19. Gennings, C.; Chinchilli, V.M.; Carter, W.H. Response Surface Analysis with Correlated Data: A Nonlinear Model Approach. *J. Am. Stat. Assoc.* **1989**, *84*, 805–809. [[CrossRef](#)]
20. Njuho, P.M.; Milliken, G.A. Analysis of Linear Models with One Factor Having Both Fixed and Random Levels. *Commun. Stat. Theory Methods* **2005**, *34*, 1979–1989. [[CrossRef](#)]
21. Njuho, P.M.; Milliken, G.A. Analysis of Linear Models with Two Factors Having Both Fixed and Random Levels. *Commun. Stat. Theory Methods* **2009**, *38*, 2348–2365. [[CrossRef](#)]
22. Kotchaporn, S.; Araveeporn, A. Modifications of Levene’s and O’Brien’s Tests for Testing the Homogeneity of Variance Based on Median and Trimmed Mean. *Thail. Stat.* **2018**, *16*, 106–128.

23. Sullivan, L.M. Repeated Measures. *Circulation* **2008**, *117*, 1238–1243. [[CrossRef](#)]
24. Armstrong, R.A. Recommendations for Analysis of Repeated-Measures Designs: Testing and Correcting for Sphericity and Use of MANOVA and Mixed Model Analysis. *Ophthalmic Physiol. Opt.* **2017**, *37*, 585–593. [[CrossRef](#)]
25. Freund, R.J.; Wilson, W.J.; Mohr, D.L. Design of Experiments. In *Statistical Methods*; Elsevier: Amsterdam, The Netherlands, 2010; pp. 521–576. [[CrossRef](#)]
26. Geisser, S.; Greenhouse, S.W. An Extension of Box's Results on the Use of the F Distribution in Multivariate Analysis. *Ann. Math. Statist.* **1958**, *29*, 885–891. [[CrossRef](#)]
27. Huynh, H.; Feldt, L.S. Estimation of the Box Correction for Degrees of Freedom from Sample Data in Randomized Block and Split-Plot Designs. *J. Educ. Stat.* **1976**, *1*, 69. [[CrossRef](#)]
28. Box, G.E.P. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *Ann. Math. Statist.* **1954**, *25*, 290–302. [[CrossRef](#)]
29. Verma, J.P. *Repeated Measures Design for Empirical Researchers*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015.
30. Conover, W.J.; Guerrero-Serrano, A.J.; Tercero-Gómez, V.G. An Update on 'a Comparative Study of Tests for Homogeneity of Variance'. *J. Stat. Comput. Simul.* **2018**, *88*, 1454–1469. [[CrossRef](#)]
31. Ott, L.; Longnecker, M. *An Introduction to Statistical Methods & Data Analysis*, 7th ed.; Cengage Learning: Boston, MA, USA, 2016.
32. Mauchly, J.W. Significance Test for Sphericity of a Normal N-Variate Distribution. *Ann. Math. Statist.* **1940**, *11*, 204–209. [[CrossRef](#)]
33. Statistical Methods and Data Analytics. UCLA: Statistical Consulting Group. Available online: <https://stats.oarc.ucla.edu/sas/seminars/sas-repeatedmeasures/> (accessed on 6 February 2022).
34. Moskowitz, D.S.; Hershberger, S.L.; American Psychological Association (Eds.) *Modeling Intraindividual Variability with Repeated Measures Data: Methods and Applications*; Multivariate Applications Book Series; L. Erlbaum Associates: Mahwah, NJ, USA, 2002.
35. Hocking, R.R. *The Analysis of Linear Models*; Brooks/Cole Pub. Co: Monterey, CA, USA, 1985.
36. Harville, D.A. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *J. Am. Stat. Assoc.* **1977**, *72*, 320–338. [[CrossRef](#)]
37. SAS Institute Inc. *SAS/STAT®14.3 User's Guide*; SAS Institute Inc.: Cary, NC, USA, 2017.
38. Milliken, G.A.; Johnson, D.E. *Analysis of Messy Data. 3: Analysis of Covariance*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2002.
39. Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* **1974**, *19*, 716–723. [[CrossRef](#)]
40. Kuehl, R.O. *Design of Experiments: Statistical Principles of Research Design and Analysis*, 2nd ed.; Brooks/Cole, Cengage Learning: Belmont, CA, USA, 2000.
41. Wolfinger, R.D.; Chang, M. Comparing the SAS GLM and Mixed Procedures for Repeated Measures. In Proceedings of the Twentieth Annual SAS Users Groups Conference; SAS Institute Inc: Cary, NC, USA, 1995.
42. Hamer, R.M. Mixed-Up Mixed Models: Things That Look Like They Should Work but Don't, and Things That Look Like They Shouldn't Work but Do. In Proceedings of the Twenty-Fifth Annual SAS®Users Group International Conference, Indianapolis, Indiana, 9–12 April 2000.
43. McLean, R.A.; Sanders, W.L.; Stroup, W.W. A Unified Approach to Mixed Linear Models. *Am. Stat.* **1991**, *45*, 54–64. [[CrossRef](#)]