EPH - International Journal of Mathematics and Statistics
ISSN (Online): 2208-2212
Volume 3 Issue 1 June 2017

DOI:https://doi.org/10.53555/eijms.v4i1.17

POWER ANALYSIS ON REPEATED MEASURES DESIGNS

Taylor King[1], Curt Doetkott[2], Rhonda Magel[3]*
*[1,2,3]*Department of Statistics North Dakota State University Fargo, ND 58108*

Corresponding Author:-
E-mail:- Rhonda.magel@ndsu.edu 701-231-7177

Abstract:-
A simulation study comparing powers of the multivariate analysis (PROC GLM) with using a mixed model (PROC MIXED) using a variety of covariance structures for various treatment, time, and interaction affect sizes in a repeated measures design is conducted. Type 1 errors are estimated. Powers are estimated for a variety of covariance structures when the actual covariance structure is AR (1). It was found that the estimated powers for treatment effect were all very similar with PROC MIXED with the correct covariance structure having the largest estimated powers. When testing for time and interaction effect, it was found when in doubt that it was better to use a simpler covariance structure. The powers were generally higher in this case than with a more complex covariance structure.

Key words: - *Variance Components; Compound Symmetry; First-Order Autoregressive; Toeplitz; Unstructured*

1. INTRODUCTION

There are many different fields of study that use repeated measures in research, such as animal science, pharmaceuticals, sports medicine, and psychology. Repeated measures experiments involve multiple subjects where measurements are taken on each subject over time or space. Often, repeated measures experiments follow a completely randomized design where the subjects are randomly assigned to different treatments. In repeated measures studies such as these, the treatment effect, time effect, and the interaction effect between treatment and time may all be of interest. A good example of a repeated measures experiment with treatment and time effects may be a pharmaceutical company testing the effects of a drug on blood pressure. The subjects of this experiment would be randomly assigned to a placebo or treatment group, and then each subject would have their blood pressure taken over a series of time intervals, generally equally spaced. In this case, the treatment effect is the between-subjects effect, while the time effect is the withinsubjects effect. The completely randomized design described above follows a mixed model, shown below in model (1).

$$yijk = \mu + \tau i + \alpha j + \gamma ij + (i) + eijk \quad (1)$$

Where y_{ijk} is the response variable measured on the k[th] subject, in the i[th] treatment group, on the j[th] time. μ Is the overall mean, τ_i is main effect for treatment, α_j is the main effect for time, and $\gamma_{ij}$is the interaction effect for treatment and time. $_{(i)}$ is the random effect for the k[th] subject with the i[th] treatment, and $d_{k(i)}$ ~ N(0, σ_S^2). $e_{ijk}$is the random error for the k[th] subject, in the i[th] treatment group, on the jth time. e_{ijk} ~ N (0, $\sigma_R{}^2$).

Model (1) can also be written in matrix form. Littell, Milliken, Stroup, Wolfinger, and Schabenberger [1] give good discussion and background on this model. The mixed model in matrix form is shown below in model (2).

$$\mathbf{Y = X\beta + Zu + e} \quad (2)$$

Where **X** is the design matrix for the fixed effects, **β** is the vector of fixed effects, **Zu** denotes the random subject effects where **Z** is the design matrix for the random effects and **u** ~ N(0, **G**), and **e** ~ N(0, **R**).

Repeated measures experiments have historically been analyzed using three different statistical methods. These methods consist of the traditional univariate analysis of variance, multivariate analysis of variance, and mixed models. SAS, a statistical software, has many built in procedures that enable researchers to perform these three statistical methods on repeated measures data [2]. SAS was released in 1976, and univariate analysis of variance was the only method available in SAS at the time to perform repeated measures analysis [3]. This analysis can be performed using PROC GLM in SAS. One issue with using the traditional univariate analysis to analyze repeated measures data is that it does not consider the possible within-subjects factor when blocking by subject is not used. The measurements taken on the same subject over time tend to be correlated, therefore the usual assumption of independent errors does not hold. For this reason, the results given by this method are generally not reliable.

In 1984, SAS improved the ability to analyze repeated measures data by adding the REPEATED statement to PROC GLM [4]. This statement allows for the use of the multivariate analysis of variance method because the repeated variable can be specified. This method can only model the data to have an unstructured covariance structure, but how severely the model assumptions may be violated can be assessed.

The last of the three methods is the mixed model method, which can be performed though PROC MIXED in SAS. The MIXED procedure was added to SAS in 1992 [3]. One advantage of the mixed model method over the univariate and multivariate methods, is that it is the only method that allows for analysis on repeated measures data that has missing data values from individual subjects. The univariate method simply ignores the missing observations, and the multivariate method simply ignores the subjects with missing observations. Additionally, the MIXED procedure allows for specification of the repeated variables, random effects variables, and the covariance structure. By choosing the covariance structure, researchers can take the withinsubjects correlations into account. However, choosing the correct covariance structure may be difficult.

There are five common covariance structures that will be discussed here [5], [6], [7],[8]. Variance Components (VC) is the simplest covariance structure because only the diagonal needs to be estimated. The diagonal elements in the structure are equal to$\sigma_k^2 * 1$, which correspond to the k parameters for this covariance structure. The off-diagonal elements are assumed to be equal to zero. This structure is essentially assuming independent errors.

$$VC = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

Compound Symmetry (CS) should be used when measures on the same subject have the same correlation. CS has two parameters, one being the subject correlation, and the other being the subject variance component, σ_s^2.

$$CS = \sigma_s^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

First-Order Autoregressive (AR(1)) can only be used when times are equally spaced. The correlation for each pair of observations that are measured n units apart is equal to ρ^n. AR(1) has two parameters.

$$AR(1) = \sigma_s^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Toeplitz (TOEP) is similar to AR(1) because the measurements that are taken at closer times have similar correlations (Kincaid, 2005). TOEP has k parameters, which is more than CS and AR(1) so it is considered more complex than CS and AR(1).

$$TOEP = \sigma_s^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

Unstructured (UN) is the most complex covariance structure because it has k(k+1)/2 parameters. These parameters have no restrictions since each variance can be different and each covariance can be different [7].

$$UN = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

1.1 Repeated Measures Analysis in SAS

As stated earlier, the univariate analysis approach was the first method offered in SAS. This approach is done through PROC GLM by putting the treatment variable in the MODEL statement and the time variable in the RANDOM statement [9]. This approach requires that the covariance structure for the data has sphericity. A covariance structure has sphericity if the diagonal variance components are all equal, and the off-diagonal covariance components are all equal. The univariate method requires this because it does not take into consideration that the errors within the subjects might be more strongly correlated over time. Compound Symmetry satisfies the sphericity requirement. Field [10] researched the possible issues that can occur when analyzing repeated measures data with a univariate approach, specifically when sphericity does not hold. Sphericity can be tested in PROC GLM using Mauchly's Test. When the test is not significant then we can assume sphericity. When the assumption of sphericity does not hold, the Greenhouse Giesser correction (GG) or the Huynh and Feldt correction (HF) can be used in PROC GLM to correct the degrees of freedom [10].

Another option when sphericity does not hold is to use the multivariate approach in PROC GLM. This was the second approach available for the analysis of repeated measures in SAS when the REPEATED statement was added to PROC GLM. For the multivariate analysis, you put the treatment variable in the MODEL statement and the time variable in the REPEATED statement [11]. The multivariate analysis in PROC GLM assumes that the repeated measures data follows an unstructured covariance structure [12]. As discussed earlier, UN is the most complex covariance structure as it has the most parameters, so if the data truly follows a simpler covariance structure then the multivariate analysis may not be the most reliable because it will have low power. Wolfinger and Chang [11] point out that in PROC GLM, the analysis on the fixed effects is the same whether the univariate method is being used or the multivariate method is being used. The univariate and multivariate analysis only differ for the within-subjects factor, time. The model is estimated using the method of moments in PROC GLM. This does not allow for any missing observations. When just one observation value within a subject is missing, then that entire subject is thrown out for PROC GLM to estimate the model.

Of the three methods considered, the most recent addition to SAS that allows for analysis of repeated measures is PROC MIXED, which can perform the mixed model method of analysis. When using PROC MIXED, the model is estimated using the REML (restricted maximum likelihood) method. This method allows for missing observations in the dataset, which is one benefit of using PROC MIXED as opposed to PROC GLM. Another advantage of using PROC MIXED to analyze repeated measures data is that it allows for specification of the covariance in the TYPE option. The trouble, however, is that picking the correct covariance structure for the data can be difficult and require background knowledge about the behavior of the various covariance structures.

The multivariate approach (PROC GLM) and the mixed models approach (PROC MIXED) have been compared by multiple researchers, with most preferring PROC MIXED, [13], [14]. Stroup [15] discusses the pros and cons of PROC MIXED, with an outline of cautions that need to be taken when using the mixed model approach.

1.2 Repeated Measures Simulation Research

There have been multiple papers published to help researchers better understand repeated measures analysis using PROC MIXED. Some of these papers describe the best way too efficiently and correctly code PROC MIXED for repeated measures analysis. When using PROC MIXED, there are two statements that can be specified, the RANDOM statement and the REPEATED statement. The TYPE option can be used in either the RANDOM of REPEATED statement to specify the covariance structure. It has been recommended to use PROC MIXED with the REPEATED statement when analyzing mixed models [9]. When only using the REPEATED statement, the TYPE option must be used in that statement to specify the covariance structure. Additionally, Kiernan, Tao, and Gibbs [16] suggest that the use of the SUBJECT= option in the REPEATED statement can improve code efficiency. Mixed model analysis can be very time intensive, so it is important to have the most efficient code possible. In addition to being time intensive, PROC MIXED can also have convergence issues. Convergence can be an issue if the estimates are approaching too closely to zero. This issue can be solved by adjusting the lower bound [5].

Choosing the correct covariance structure is another topic that has been researched for the mixed model method of analysis for repeated measures. Littell et al. [9] states that estimating the covariance structure for the data is one of the first steps in the analysis. Graphical tools to visualize the covariance structure are suggested, as well as looking at the various information criteria to determine the covariance structure with the best fit. A good example of estimating the covariance structure for repeated measures data can be seen by Littell, Pendergast, and Naterajan [17]. It has also been suggested by researchers to use UN as the covariance structure when performing mixed model analysis on repeated measures data, especially when unsure which structure fits best [6], [17]. Although using UN has been suggested, it has also been determined that UN is not a suitable covariance structure when there is a small number of subjects compared to the number of repeated measures [18]. Additionally, UN has many parameters, so loss of power can be expected.

Along with literature regarding how to use SAS PROC MIXED, and determining the correct covariance structure for the analysis, there are also various simulation studies that have been conducted. Most of these simulations studies are done to determine the control over Type I error with repeated measures data under various distributions with different means and variances, as well as with various sample sizes and correlations.

Oberfeld & Franke [19] ran simulations to determine Type I error control for repeated measures data with small sample size and non-normality in the response variable. They determined that the tests to analyze repeated measures are not appropriate to use when the data is not normal. There has also been research on controlling Type I error and power analysis using the mixed model Satherthwaite F Test and the multivariate Welch-Test [20]. Keselman et al. [20] were not able to say that one test controlled Type I error and had stronger power better than the other, which still leaves the question about which approach to use.

Kowalchuk et al. [7] simulated data with between-subjects effect and within-subjects effect with multiple different covariance structures, as well as with unequal group sizes. From their study, they concluded that the KR adjusted degrees of freedom control the otherwise inflated Type I errors of complex covariance structures, such as UN. Note that the KR adjusted degrees of freedom estimate the denominator degrees of freedom for the F test to allow for inference of the fixed effects when sample size is small [21]. Kowalchuk et al. [7] also discussed the importance that the simulated Type I error estimates should stay within a certain interval about alpha. Two possible intervals were mentioned in their paper – the binomial standard error interval, and the liberal criterion of robustness. If using the binomial standard error to calculate this interval, then the Type I error values should stay within $2 * \left[\frac{\propto(1-\propto)}{N}\right]^{1/2}$, where N is the total number of samples. The liberal criterion of robustness that states the Type 1 error estimates should stay within 0.5α to 1.5α in ordered to be considered controlled [22]. For the simulation study conducted in this paper, these intervals were considered when checking the Type I error estimates.

Guerin and Stroup [5] also conducted a simulation study to determine the Type I error for repeated measured data. They too found that using the KR adjustment in PROC MIXED can help avoid the overly inflated Type I errors that result from models with complex covariance structures. Additionally, Guerin and Stroup [5] analyzed four model selection criteria: Akaike Information Criteria (AIC), Schwarz Bayesian Information Criteria (BIC), Corrected Akaike Information Criteria (AICC), and Hanan & Quinn Information Criteria (HQIC) to determine which criteria led to choosing the correct covariance structure most often. Guerin and Stroup [5] found that the AIC is the best criterion to use when choosing a model because it chooses a slightly more complex model, but not too complex of a model that power could be greatly impacted.

In this present study, these issues were explored more in depth through repeated measures data sets that were simulated and then tested. Simulations were used to create many samples of repeated measures data with a specific covariance structure and various simulation parameters such as sample sizes, number of repeated measures, effect sizes, and correlations. Then, this data was analyzed using PROC GLM with the REPEATED statement for a multivariate analysis of variance method, as well as with the PROC MIXED to perform the mixed model method. These results were then compared to determine which method performs better. The data was also analyzed after specifying various incorrect covariance structures in PROC MIXED to determine the consequences on power. These comparisons were made on data simulated from multiple samples sizes. This study represents an extension of a study done by Guerin and Stroup [5] who used simulations to investigate Type I error and selection criteria of covariance structures for repeated measures. Simply stated, the goals of this study are to answer the following questions:

- How does the performance of the multivariate analysis of variance method compare to the performance of the mixed model method with respect to power and Type I error?
- How does choosing the incorrect covariance structure affect the power?
  Park, Park and Davis [23] do some simulation comparisons with the multivariate analysis of variance method and the mixed model method when selecting the correct covariance structure and the incorrect covariance structure. Their work is done with sample sizes of 30 or greater. In this research, we would like to consider samples sizes of 30 or smaller. We would also like to break our analysis down into testing for treatment effects, testing for time effects, and testing for interaction effects.

2. Materials and Methods

2.1 Data Simulation

The simulation study performed in this paper involved simulating data from the marginal distribution of the mixed model shown in model (2) assuming a completely randomized design with equally spaced time intervals. Every simulation was run with 5000 samples, 2 treatment groups, an AR (1) covariance matrix, a residual variance component (σ_R^2) equal to 1, a subject variance component (σ_S^2) equal to 4, and a mean equal to 5. The changing simulation parameters consisted of a various number of sample sizes, a various number repeated measurements, various effect sizes, and different correlations for the AR (1) covariance matrix.

The combinations of these different parameters, along with how the power for each simulation was attained, will be discussed next in section 3.2. First, a detailed walk through of how the data was simulated will be discussed and shown with an example. One of the simulations run in this study used data that was generated with 5 subjects per treatment group, 4 repeated measurements, and correlation equal to 0.25. This specific simulation is used in the following paragraphs to show examples of the generated data, and can be visualized in Figure 1.

Subject Time Trt	1 1	2 1	3 1	4 1	5 1	6 2	7 2	8 2	9 2	10 2
1										
2										
3										
4										

Figure 1. Repeated Measures Experiment when Number of Subjects equals 10

The marginal distribution of model (2) is such that $\mathbf{Y}\sim N(\mathbf{X\beta}, \mathbf{V})$ where $\mathbf{V}=\mathbf{ZGZ}^T+\mathbf{R}$ [1]. There are four main parts that had to be specified to obtain **Y**. The first part consisted of generating the variance. Littell et al. [1] describe that one way of obtaining **V** is to get rid of **Z** and **G** and set **R** to be a block matrix with each block corresponding to a subject. This approach was taken in this study. Therefore, our initial covariance matrix (**R₀**) was as follows,

$$\mathbf{R_0}=\begin{bmatrix} 1 & 0.25 & 0.0625 & 0.015625 \\ 0.25 & 1 & 0.25 & 0.0625 \\ 0.0625 & 0.25 & 1 & 0.25 \\ 0.015625 & 0.0625 & 0.25 & 1 \end{bmatrix}.$$

From here, this matrix was used as the block matrix for each subject, and then the block matrix was multiplied by the subject variance component, σ_S^2, which was set to equal 4 for this simulation study. The final **R** matrix, which represents the Var(**Y**), is then equal to 40 by 40 matrix to correspond to each of the 40 observations for each sample. ((2)(5)(4) = 40 rows because our example simulation had 2 treatment groups, 5 subjects per group, and 4 repeated measurements per subject.)

This **R** matrix was next used to obtain the subject random error. The subject random error was generated to follow a normal distribution with a mean of zero, and a variance of **R**. For this simulation, the subject error matrix had 40 rows corresponding to each observation, and 5000 columns corresponding to each sample.

The third part that had to be specified is $\mathbf{X\beta}$, where **X** is the design matrix and $\boldsymbol{\beta}$ is the vector of parameters. The vector $\boldsymbol{\beta}$ was set such that $\boldsymbol{\beta}^T = (\mu, \tau_1, \tau_2, \alpha_{1,2}, \alpha_3, \alpha_4) = (5, 0, 0, 0, 0, 0, 0)$ when there is no effect size being simulated. The design matrix is then simulated to have 7 columns that correspond to each parameter in $\boldsymbol{\beta}$, and 40 rows which correspond to each observation. Therefore, the design matrix is as follows.

$$X=\begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This $\mathbf{X\beta}$ portion is then repeated 5000 times because that was the number of samples being simulated.

Lastly, the random error components were generated to follow a normal distribution with a mean of zero, and a variance of σ_R^2, where σ_R^2 was set to equal 1. This vector was repeated 5000 times, and therefore was a matrix with 40 rows corresponding to each observation, and 5000 columns corresponding to each sample.

Y was then obtained by adding **Xβ**, the subject random error components, and the random error components. The described steps above were done using PROC IML in SAS.

2.2 Simulation Parameters and Finding Power

Like the study done by Guerin and Stroup [5], there were multiple simulations run with certain parameters set to equal various values for this study. All simulations were run with 2 treatment groups, 5000 samples, and an AR (1) covariance structure. The values used for the simulation parameters are as listed below.

- Total number of subjects: s = 10, 20, 30
- Number of repeated measures: k = 4, 6
- Correlation used in AR(1) covariance structure: ρ = 0, 0.25, 0.75
- Effect size: effect size on treatment only with (½)σ_S, σ_S, 2σ_S; effect size on time only with (½)σ_S, σ_S, 2σ_S; effect size on interaction term only (½)σ_S, σ_S, 2σ_S

From the diagonal of **R**, we know that σ_S^2= 4. Therefore (½) = 1, σ_S = 2, and 2σ_S = 4. The effect sizes are specified in **β**. For example, when specifying an effect size of σ_S on the treatment effect only and k = 4, then β^T = (5, 0, 2, 0, 0, 0, 0). When specifying an effect size of σ_S on the treatment effect only and k = 6, then β^T = (5, 0, 2, 0, 0, 0, 0, 0, 0). To generate realistic data when simulating an effect size on time, each time was increased by a constant amount. For example, for an effect size of σ_S^2 on the time effect when k = 4, β^T = (5, 0, 0, 0, 0.67, 1.33, 2). For an effect size of σ_S^2 on the time effect when k = 6, β^T = (5, 0, 0, 0, 0.4, 0.8, 1.2, 1.6, 2). The data were simulated using all combinations of s, k, and ρ for each effect size. This resulted in 18 simulated datasets per effect size, therefore a total of 162 simulated data sets were run in this study to analyze power.

Each of these data sets was then analyzed using PROC GLM to obtain the power for multivariate analysis method, as well as analyzed using PROC MIXED to obtain the power for analysis of the mixed model method. Example code for PROC GLM and PROC MIXED can be seen below in Figures 2 and 3, respectively. Note that in PROC MIXED the KR adjusted degrees of freedom was used. This is because the Type I error was not controlled for all covariance structures without it. Also, Kenward & Roger [21] suggest to use their approach when samples are small, which is the case for this simulation study.

```
Proc mixed data= rm.uv_AR1_4_10; by sample; class trt time
        subj_id; model y = trt | time / ddfm=kr; repeated time /
        subject=subj_id type=AR (1);
title2 "Repeated Measures ANOVA using Mixed Model Approach --
        AR (1)"; run;
```

Figure 2. Example of PROC MIXED code when TYPE = AR (1)

As shown, both PROC GLM and PROC MIXED are set up to run the tests by sample. From PROC GLM, the Wilk's Lambda p-value was obtained for each sample when testing for the time effect or treatment*time effect. The number of samples with p-values less than or equal to 0.05 was divided by the total number of samples to determine power. The power for testing for the treatment effect size in PROC GLM was found by taking the number samples with Type III Sum of Squares ANOVA p-values less than or equal to 0.05 and dividing by the total number of samples

```
Proc glm data=rm.mv_AR1_4_10; by sample;
        Class trt;
        Model y1 – y4 = trt / nouni;        repeated
time 4;
            title2 'Repeated Measures ANOVA for Effect of Time on
        Response';
        Run;
```

Figure 3. Example of PROC GLM code with 4 Repeated Measures

In PROC MIXED, the significance for treatment, time, or treatment*time can all be found by looking at the p-values obtained from the F statistics in the Type III Sum of Squares ANOVA. Like with PROC GLM, the power was obtained by taking the number of samples with p-values less than or equal to 0.05, and dividing by the total number of samples.

From these power results, the trend in power was observed for each analysis which was used to try to determine which analysis is superior. Additionally, when the data sets were run in PROC MIXED, they were run with the TYPE option set to be equal to VC, CS, AR(1), TOEP, and UN. This was done to determine how choosing the incorrect covariance structure can impact the power.

Simulations were run with different combinations of sample sizes, number of repeated measures, and different correlations. These simulated data sets were then analyzed to determine the power for both the multivariate and mixed model methods of analysis. Type 1 error rates were estimated. Powers were also estimated for different treatment effect sizes, different time effect sizes, and different interaction effect sizes.

3. Results

3.1 Checking Type 1 Error Rates

In all situations considered, we first estimated the type one error rates based on a stated type 1 error rate of 0.05. In these cases, we also estimated the type 1 error rates for the univariate analysis of variance (PROC GLM UV). Table 1 displays the Type I error rates for the treatment effect for all the situations considered. Specified Type 1 error rates were not maintained in the univariate analysis of variance case (PROC GLM UV) of when the Variance Component structure was

used in PROC MIXED when the actual variance component was AR(1). The multivariate method has the same between-subjects effect test as the univariate method does when a compound symmetry covariance structure is used. For the other methods and specified covariance structures, the Type I error rates were kept within appropriate limits of alpha equal to 0.05. Note that the PROC MIXED analysis with CS specified as the covariance structure resulted in Type I error rates equal to the PROC GLM F test since the PROC GLM assumes CS as the covariance structure when testing the between-subjects effect. The Type 1 Error rates for time effect and interaction effect were within the appropriate limits when alpha was equal to 0.05, or they were really conservative as was the case for the univariate analysis and PROC MIXED VC. As the correlation increased, the Type I error rates dramatically decreased, in these two cases.
Overall, the PROC GLM multivariate method analyses displayed controlled Type I error rates. The PROC MIXED analyses showed controlled Type I error for all covariance structures specified excluding the VC covariance structure. Therefore, the VC covariance structure was not analyzed further in the power analysis. The univariate analysis methods were not of interest in this study, but the Type I error results were included in this section to provide some helpful insight. The inflation in the Type I error rates for the PROC GLM univariate analyses with no blocking results show why caution should be used when assuming all observations of a repeated measures experiment are independent. We will only consider the power analysis for the PROC GLM multivariate method, and the PROC MIXED method with CS, TOEP, UN, or AR(1) as the covariance structure.

Table 1. Type I Error Rates for Treatment Effect

Number of Samples = 5,000 Binomial Proportions CI: (0.0438, 0.0562)							
Method of Analysis	Sample Size	rho = 0 k = 4	k = 6	rho = 0.25 k = 4	k = 6	rho = 0.75 k = 4	k = 6
PROC GLM F test	10	4.86	4.82	4.94	5.16	4.84	4.38
	20	4.76	5.06	5.38	5.24	4.78	5.16
	30	5.10	4.74	5.04	4.50	5.52	4.70
PROC GLM UV No Blocking	10	4.58	4.66	8.96	9.80	21.42	26.72
	20	5.04	4.76	9.46	10.08	21.44	26.94
	30	5.00	5.02	8.76	9.42	22.40	27.50
PROC MIXED VC	10	4.58	4.66	8.96	9.80	21.42	26.72
	20	5.04	4.76	9.46	10.08	21.44	26.94
	30	5.00	5.02	8.76	9.42	22.40	27.50
PROC MIXED CS	10	4.86	4.82	4.94	5.16	4.88	4.38
	20	4.76	5.06	5.38	5.24	4.78	5.16
	30	5.10	4.74	5.04	4.50	5.52	4.70
PROC MIXED TOEP	10	4.76	4.92	4.88	5.42	4.84	4.76
	20	4.82	5.06	5.34	5.10	4.88	5.30
	30	5.08	4.78	5.02	4.58	5.44	4.82
PROC MIXED UN	10	4.86	4.82	4.94	5.16	4.88	4.38
	20	4.76	5.06	5.38	5.24	4.78	5.16
	30	5.10	4.74	5.04	4.50	5.52	4.70
PROC MIXED AR(1)	10	4.74	4.62	5.04	5.62	6.20	6.78
	20	4.80	5.16	5.52	5.18	5.74	6.88
	30	5.30	4.86	5.10	5.02	6.38	6.44

3.2 Powers for Treatment Effect

The Table 2 display the findings for the power of detecting an effect size of σ on the treatment effect. Power results from the other two effect sizes are similar, as far as which procedures have the higher powers. The PROC MIXED AR (1) analysis does show slightly higher power, although not drastically. Recall, AR (1) is the correct covariance structure. The power of PROC GLM is similar to the power for PROC MIXED when CS, TOEP, and UN are used and only a little smaller when AR(1) is used. It can be noticed that as the number of repeated measures increases, the power also increases. It can also be noticed that as the correlation increases, the power decreases.

Table 2. Power for Detecting an Effect Size of σ_S for Treatment Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0 k = 4	k = 6	rho = 0.25 k = 4	k = 6	rho = 0.75 k = 4	k = 6
PROC GLM UV F test	10	70.32	86.36	58.04	72.16	36.14	41.10
	20	96.08	99.58	90.34	97.64	66.20	73.92
	30	99.74	100	98.34	99.78	85.40	90.34
PROC MIXED CS	10	70.32	86.36	58.04	72.16	36.14	41.10
	20	96.08	99.58	90.34	97.64	66.20	73.92
	30	99.74	100	98.34	99.78	85.40	90.34
PROC MIXED TOEP	10	69.68	85.88	57.98	71.62	36.76	41.52
	20	95.98	99.60	90.30	97.52	66.30	73.94
	30	99.74	100	98.26	99.76	85.48	90.62
PROC MIXED UN	10	70.32	86.36	58.04	72.16	36.14	41.10
	20	96.08	99.58	90.34	97.64	66.20	73.92
	30	99.74	100	98.34	99.78	85.40	90.34
PROC MIXED AR(1)	10	74.04	91.18	61.66	77.92	40.46	48.92
	20	96.90	99.72	91.48	98.06	69.02	78.96
	30	99.78	100	98.54	99.84	87.08	92.96

3.3 Powers for Time Effect

The power results for detecting the effect size of σ on the time effect can be seen on Table 3. In this case, the power did not increase as the number of repeated measures increased. Instead, the power was higher for 4 repeated measures than it was for 6 repeated measures. One possible explanation for this may be that the effect size is spread across more time periods when k = 6, which means the effect is incrementing by smaller amounts. This may impact how well the effect size is detected for the time effect.

There was not a consistent trend found as correlation increased that fit all covariance structures and methods. The PROC MIXED with CS as the covariance structure was the only test with power results that consistently increased as the correlation increased. This could possibly be due to the simplicity of the model, and the small sample sizes tested in this study. Most methods and covariance structures seemed to have a trend where the power would be higher for a correlation of 0 and 0.75, and a lower power when the correlation was 0.25. One possible explanation for the larger power when the correlation is equal to 0, is that the covariance structures are not necessarily needed to estimate the model since a correlation of 0 implies independence.

The power for the PROC MIXED results when UN is the specified covariance structure produced the lowest power trends compared to the other covariance structures. Notice these results are the same as the multivariate Wilk's Lambda power, which is due to the KR degrees of freedom applied in the PROC MIXED procedure. Recall that the PROC GLM multivariate method assumes the data has an unstructured covariance structure which also explains the similar results. The unstructured covariance structure has a large number of parameters that need to be estimated which can explain why it resulted in the lowest power. The PROC MIXED power results when CS is the specified covariance structure produced the highest power compared to the other covariance structures. Compound symmetry is one of the simpler structures that was tested which may explain why it resulted in the highest power.

Table 3. Power for Detecting an Effect Size of σ_S for Time Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25 k		rho = 0.75 k	
		k = 4	k = 6	= 4	k = 6	= 4	k = 6
PROC GLM Wilk's Lambda	10	22.94	16.88	20.72	13.72	30.00	14.14
	20	58.88	54.20	56.42	46.78	73.34	50.20
	30	82.86	82.62	79.46	73.30	93.04	79.16
PROC MIXED CS	10	33.64	37.18	37.16	38.46	58.74	55.12
	20	67.18	70.06	70.10	71.56	89.62	85.10
	30	87.34	89.64	88.16	88.18	98.42	95.90
PROC MIXED TOEP	10	30.12	31.86	28.38	26.30	39.12	29.60
	20	64.46	65.74	61.82	58.04	77.78	61.36
	30	86.10	87.62	82.50	79.94	94.90	84.06
PROC MIXED UN	10	22.94	16.88	20.72	13.70	30.00	14.14
	20	58.88	54.20	56.42	46.78	73.34	50.20
	30	82.86	82.62	79.46	73.30	93.04	79.16
PROC MIXED AR(1)	10	33.00	36.70	30.20	29.66	34.62	26.68
	20	67.36	69.36	63.58	61.24	73.20	58.34
	30	87.30	89.84	84.20	81.84	92.84	80.64

3.4 Powers for Interaction Effect

The power results for detecting the effect size of   on the interaction effect can be seen in effect can be seen in Table 4. CS is again the only covariance structure where an increase in correlation always resulted in an increase of power. Results overall are similar to the results of power for time effect.

Table 4. Power for Detecting an Effect Size of σ_S for Interaction Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	23.84	15.66	21.48	13.74	30.22	15.18
	20	58.40	55.82	54.68	45.80	74.72	51.58
	30	82.36	82.24	79.52	73.80	93.60	78.82
PROC MIXED CS	10	34.66	36.40	37.34	38.68	59.24	54.04
	20	66.48	71.96	69.48	71.14	89.76	85.36
	30	86.60	89.46	87.46	88.78	98.40	96.22
PROC MIXED TOEP	10	32.38	31.72	28.88	27.20	40.02	27.96
	20	63.50	67.40	60.46	58.08	78.30	61.04
	30	85.22	87.20	82.30	80.42	94.88	84.20
PROC MIXED UN	10	23.84	15.66	21.48	13.74	30.22	15.18
	20	58.40	55.82	54.68	45.80	74.72	51.58
	30	82.36	82.24	79.52	73.80	93.60	78.82
PROC MIXED AR(1)	10	34.06	36.38	30.84	29.56	35.32	25.30
	20	66.40	71.30	62.96	61.58	74.04	57.24
	30	86.70	89.62	83.54	83.02	92.60	81.10

4. Discussion

There are many disciplines that use repeated measures when conducting their studies. Past research on repeated measures has been mostly related to Type I error rates, or how to conduct various methods of analysis, or determining the correct

covariance structure. This study focused on the power analysis on repeated measures data. The study investigated how the performance of multivariate analysis compares to the performance of the mixed model method. With respect to the between-subjects effect, treatment, the multivariate method with PROC GLM had reasonable power results. The PROC MIXED power analysis with AR (1) resulted in the highest power, which was expected because it was the true covariance structure. The multivariate method did not perform better than the power analysis with the true covariance structure specified, but did perform just as well as the mixed models approach when specifying the incorrect covariance structure.
With respect to the within-subjects effect, time, the multivariate method may not be preferred. When trying to detect a small effect size, the multivariate method seemed to perform similarly to the mixed model method. Although, when the effect size was large, the power for the multivariate method did not increase as much as the power did for some of the covariance structures specified for the mixed model. One can see that when the effect size is σ_S as given in Table 4, the powers for the multivariate method are smaller. This could be due to the multivariate method assuming an unstructured covariance structure. An unstructured covariance structure is more complex than the true covariance structure for the data, and therefore could be causing the low power. Overall, if testing for the effect on the within-subjects effect is of interest, the multivariate method may not be the best approach if the true data has a simpler covariance structure. The multivariate results were the same as the results from the mixed model method when UN was the specified covariance structure, and therefore choosing a covariance structure other than UN when interested in the within-subjects effect may also be preferred when the data truly has a simpler covariance structure and the sample size is small.
The study also investigated how choosing the incorrect covariance structure affected the power. As just discussed, if researchers are only concerned about the between-subjects effect, then choosing the incorrect covariance structure may only have a slight impact on power. Choosing the correct covariance structure in this simulation study did result in the highest power, but if a researcher can get close to the correct covariance structure then the power should not be greatly affected. Choosing the incorrect covariance structure when testing for an effect on the withinsubjects effect may be more important. The power analysis for the correct covariance structure, PROC MIXED with AR (1) specified as the covariance structure, did not result in the highest power for detecting an effect size in the within-subjects effect. The covariance structure with the lowest power was the PROC MIXED power analysis with UN as the specified covariance structure. The covariance structure with the highest power was the PROC MIXED power analysis with CS as the specified covariance structure. This is not surprising as unstructured is a complex covariance structure, so a low power is expected when sample sizes are small. Compound symmetry is one of the simplest covariance structures. Therefore, if not sure which error structure to choose, going with the slightly simpler choice may give a researcher more power to find a difference in the within-subjects effect.

5. Conclusion

When testing for treatment effect differences, powers are similar between the multivariate analysis (UN) and the mixed models approach. When testing for time effect and interaction effect, the mixed model approach gave higher powers except when the unstructured covariance structure was used. Overall, the findings support that choosing the slightly simpler covariance structure may be preferred over choosing the more complex covariance structures because complex covariance structures have more parameters to estimate and can therefore reduce power. The research conducted in this paper could be continued on in many ways to obtain more knowledge about the power of repeated measures. Simulating data with different ratios of subject variance components to residual variance components and then determining power trends would be a good continuation of this study. Additionally, conducting a power analysis on repeated measures data that is unbalanced, or follows a non-normal distribution are also future research suggestions.

REFERENCES

[1]. Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). Appendix 1: Linear Mixed Model Theory. *SAS® for Mixed Models, Second Edition* (pp. 734-756). Cary, NC: SAS Institute.

[2]. SAS Institute Inc. (2017). *SAS/STAT® 14.3 User's Guide*. Cary, NC: SAS Institute Inc. Retrieved from http://support.sas.com/documentation/onlinedoc/stat/143/statug.pdf

[3]. Littell, R. C. (2011). The Evolution of Linear Models in SAS: A Personal Perspective. *SAS Global Forum 2011*. Retrieved from http://support.sas.com/resources/papers/proceedings11/325-2011.pdf

[4]. Littell, R.C., Henry, P.R., & Ammerman, C.J. (1998). Statistical Analysis of Repeated Measures Data Using SAS Procedures. *J. Animal Science, 76*. Retrieved from http://www.stat.ncsu.edu/people/arellano/courses/ST524/Fall08/Homeworks/Homework7/articles/Littell_mixed_JAS.pdf

[5]. Guerin, L. & Stroup, W. W. (2000). A Simulation Study to Evaluate PROC MIXED Analysis of Repeated Measures Data. *Annual Conference on Applied Statistics in Agriculture*. Retrieved from http://newprairiepress.org/agstatconference/2000/proceedings/15

[6]. Kincaid, C. (2005). Guidelines for Selecting the Covariance Structure in Mixed Model Analysis. *SUGI 30 Proceedings.* Retrieved from http://www2.sas.com/proceedings/sugi30/198-30.pdf

[7]. Kowalchuk, R. K., Keselman, H. J., Algina, J., & Wolfinger, R. D. (2004). The Analysis of Repeated Measurements with Mixed-Model Adjusted F Tests. *Educational and Psychological Measurement, 64.* Retrieved from http://journals.sagepub.com/doi/abs/10.1177/0013164403260196

[8]. Wolfinger, R. D. (1996). Heterogeneous Variance: Covariance Structures for Repeated Measures. *Journal of Agricultural, Biological, and Environmental Statistics, 1(2).*

[9]. Retrieved from https://www.researchgate.net/profile/Russ_Wolfinger2/publication/272579574_Heterogeneous_Variance_Covariance_Structures_for_Repeated_Measures/links/5756da6808ae5c654903f3fe/Heterogeneous-Variance-Covariance-Structures-for-Repeated-Measures.pdf
[10]. Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006a). Chapter 5: Analysis of Repeated Measures Data. *SAS® for Mixed Models, Second Edition* (pp. 160-186). Cary, NC: SAS Institute.
[11]. Field, A. (2010). A Bluffer's Guide to ... Sphericity. *BPS-MSC Newsletter 6 (1)*. Retrieved from http://www.discoveringstatistics.com/docs/sphericity.pdf
[12]. Wolfinger, R. D., & Chang, M. (1995). Comparing the SAS GLM and MIXED Procedures for Repeated Measures. *SUGI Proceedings.* Retrieved from https://support.sas.com/rnd/app/stat/papers/mixedglm.pdf
[13]. Everitt, B.S. (1995). The Analysis of Repeated Measures: A Practical Review with Examples. *Journal of the Royal Statistical Society. Series D (The Statistician), 44 (1)*. Retrieved from http://www.jstor.org/stable/2348622
[14]. Gueorguieva, R. & Krystal, J. H. (2004). Move Over ANOVA: Progress in Analyzing Repeated-Measures Data and Its Reflection in Papers Published in the Archives of General Psychiatry. *Archives of general psychiatry, 61*. Retrieved from https://www.researchgate.net/profile/RalitzaGueorguieva/publication/5412094_Move_Over_ANOVA_Progress_in_Analyzing_Repeated-
[15]. Measures_Data_andIts_Reflection_in_Papers_Published_in_the_Archives_of_General_Psychiatry/links/5617793308ae40a7199a94ca/Move-Over-ANOVA-Progress-inAnalyzing-Repeated-Measures-Data-andIts-Reflection-in-Papers-Published-in-theArchives-of-General-Psychiatry.pdf
[16]. Wang, Z. & Goonewardene, L. A. (2004). The use of MIXED models in the analysis of animal experiments with repeated measures data. *Canadian Journal of Animal Science*. Retrieved from https://era.library.ualberta.ca/files/crb68xb84k/cjas_84 (1)
[17]. Stroup, W. W. (1999). On Using PROC MIXED for Longitudinal Data. *Annual Conference on Applied Statistics in Agriculture*. Retrieved from http://newprairiepress.org/agstatconference/1999/proceedings/5
[18]. Kiernan, K., Tao, J., & Gibbs, P. (2012). Tips and Strategies for Mixed Modeling with SAS/STAT® Procedures. *SAS Global Forum 2012*. Retrieved from http://support.sas.com/resources/papers/proceedings12/332-2012.pdf
[19]. Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine, 19(13)*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/1097-0258(20000715)19:13%3C1793::AIDSIM482%3E3.0.CO;2-Q/abstract
[20]. Moser, E. B. (2004). Repeated Measures Modeling With PROC MIXED. *SUGI 29 Proceedings.* Retrieved from http://www2.sas.com/proceedings/sugi29/188-29.pdf
[21]. Oberfeld, D. & Franke, T. (2013) Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. *Behav Res, 45*. Retrieved from https://www.staff.uni-mainz.de/oberfeld/downloads/oberfeld_franke_2013_BRM.pdf
[22]. Keselman, H. J., Algina, J., Kowalchuk, R. K. & Wolfinger, R. D. (1999). The Analysis of Repeated Measurements: A Comparison of Mixed-Model Satterthwaite F Tests and a Nonpooled Adjusted Degrees of Freedom Multivariate Test. *Communications in Statistics Theory and Methods, 28(12)*. Retrieved from http://home.cc.umanitoba.ca/~kesel/cis1999b.pdf
[23]. Kenward, M. G. & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics, 53(3)*. Retrieved from http://www.public.iastate.edu/~dnett/S511/KR.pdf
[24]. Bradley, J. V. (1978). Robustness? *British Journal of Mathematical & Statistical Psychology, 31*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.2044-8317.1978.tb00581.x/abstract
[25]. Park, T., Park, J., and Davis, C. (2001). Effects of covariance model assumptions on hypothesis tests for repeated measurements: analysis of ovarian hormone data and pituitary-pteryomaxillary distance data. *Statistics in Medicine*, 20: 2441-2453.