**Two-Way Repeated Measures Mixed ANOVA: Insights for Medical Researchers Using Simulations**

**ABSTRACT:**

Two-way repeated measures Analysis of Variance (RMANOVA) is widely utilized in clinical and medical research to analyze longitudinal data when subjects are measured multiple times across different conditions. This statistical approach is particularly valuable when examining the effects of interventions over time or comparing treatment responses between groups. However, its application requires careful consideration of several statistical assumptions that, when violated, can compromise the validity of results. This paper provides a comprehensive overview of two-way repeated measures mixed ANOVA, detailing its methodological foundations, essential assumptions, implementation steps, and interpretation guidelines. Using a simulated dataset, we demonstrate the complete workflow from preliminary data assessment through analysis and results interpretation. Additionally, we provide practical implementation guidelines using common statistical software SPSS.

**KEYWORDS:** Two-way Repeated Measures ANOVA; Longitudinal Studies; Clinical Trials; Sphericity; Statistical Analysis; Interaction Effects.

**INTRODUCTION:**

Longitudinal studies in medical and clinical research frequently involve multiple measurements from the same subjects over time. Such repeated measures designs offer several advantages, including increased statistical power and the ability to track individual-level changes. Two-way repeated measures ANOVA has emerged as a powerful analytical tool for such designs, particularly when researchers need to simultaneously assess the effects of time (within-subjects factor- same subjects measured multiple times) and group assignment (between-subjects factor-different groups exposed to different conditions)1.

The two-way RMANOVA allows researchers to test three distinct hypotheses: (1) the main effect of time, (2) the main effect of group, and (3) the interaction between time and group. This interaction effect is particularly informative in intervention studies, as it indicates whether groups respond differently over time – often the primary research question in clinical trials.

Despite its widespread use, the application of two-way RMANOVA presents several challenges. The technique requires specific assumptions including normality of residuals, homogeneity of variances, and sphericity. When these assumptions are violated, alternative approaches or corrections must be employed to ensure valid statistical inference.

This paper aims to provide clinical researchers with a comprehensive guide to implementing and interpreting two-way RMANOVA. We address common challenges in study design, data preparation, assumption testing, analysis execution, and results interpretation. Through a simulated case study on nutrition awareness program, we demonstrate the complete analytical workflow and provide practical guidelines for implementation using SPSS.

**MATERIALS AND METHODS:**

**Assumptions of RMANOVA**

The key assumptions of repeated measures ANOVA are:

* *Normality* – The dependent variable is normally distributed at each time point.
* *Sphericity* – Variances of differences between time points are equal (tested by Mauchly’s test).
* *Independence* – Observations are independent across participants.
* *No Extreme Outliers* – Checked using boxplots or residual analysis.

When the assumptions of two-way repeated measures ANOVA are violated, researchers may consider several alternative approaches:

Data Transformation: When normality is violated, appropriate transformations (e.g., log, square root) may normalize the data4.

Mixed-Effects Models: Linear mixed models offer a more flexible approach that can accommodate missing data, time-varying covariates, and different covariance structures.2,3

Generalized Estimating Equations (GEE): GEE provides a robust alternative when data is skewed or when other assumptions are violated5,6.

**Sample size**

The equation for sample size calculation for the two-way repeated measures ANOVA7,8,9 is provided below.

Were,

σ=pooled (average) standard deviation of two groups, =clinically significant difference

, - Normal distribution table values, m is the number of follow-up repeated measures, ρ intraclass correlation coefficient (ICC), representing the correlation between repeated measurements (pre/post correlation).

If loss to follow-up is anticipated, the sample size can be adjusted accordingly. For instance, if a 5% dropout rate is expected, the required sample size should be divided by 0.95 (i.e., 1 - 0.05 = 0.95) to account for the anticipated loss.

**Statistical analysis**

Before conducting the two-way repeated measures ANOVA, preliminary analyses should be performed to ensure that key assumptions are met. Normality of residuals should be evaluated using the Kolmogorov-Smirnov (KS) test or the Shapiro-Wilk test. If the p-value is greater than 0.05, the data can be considered normally distributed. To assess sphericity, Mauchly’s test should be performed. If the assumption is violated (i.e., p < 0.05), the Greenhouse-Geisser correction should be applied. Descriptive statistics, such as mean and standard deviation, should be used to summarize the data at each time point for each group. Additionally, graphical representations, such as line graphs (profile plot) with error bars (95% confidence intervals), should be generated to visualize trends over time for each group.

***Post-hoc Analyses (pairwise comparison):*** After finding a significant interaction or main effect in a Two-Way Repeated Measures Mixed ANOVA, post hoc tests help identify which specific time points (or conditions) differ within a group.

* Bonferroni correction: Conservative but simple, divides α by number of comparisons
* Holm-Bonferroni: More powerful step-down procedure
* Sidak correction: Similar to Bonferroni but less conservative

***Visual representation:*** Profile plots can be used for the visual comparison of the average outcome across groups.

***Accounting for Lost follow up:*** Loss to follow up in repeated measures ANOVA (RM-ANOVA) occurs when participants miss data collection at certain time points, affecting statistical power and validity. Common handling methods include listwise deletion (removes missing cases), LOCF (carries forward the last observation), and mean imputation (replaces with group mean). More advanced approaches like linear mixed models (LMMs) and multiple imputation are preferred, as they handle missing data flexibly without excluding participants.

**ILLUSTRATION USING SIMULATED SAMPLE**

**Data Description:** To illustrate the application of a two-way repeated measures ANOVA, we simulated a dataset representing a study comparing the average nutritional knowledge levels between two groups over time. The study aims to assess the effectiveness of an educational intervention in improving knowledge of nutrition, particularly focusing on iron-rich foods and anemia prevention. The outcome variable is the nutritional knowledge score, measured on a standardized scale from 0 to 50.

The dataset includes two groups: an intervention group that received structured training on nutrition and anemia prevention and a control group that did not receive any intervention. Knowledge levels were measured at three time points: baseline (Pretest), 1 month post-intervention (Post test 1), and 3 months post-intervention (Post test 2).

Variables in the dataset include:

* Participant ID (unique identifier for each individual)
* Group (intervention or control)
* Time points: baseline (*Pre-test*), 1 month (*Post 1*), 3 months (*Post 2*)
* Nutritional Knowledge Score (0-50 scale representing knowledge level in nutrition and anemia prevention)

For illustration purposes, we considered 80 participants, with 40 in the intervention group and40 in the control group.

**Statistical Analysis (*for simulated data*):** For the simulated data, the Shapiro-Wilk test assessed normality of residuals, while Levene’s test evaluated homogeneity of variance across groups at each time point. Mauchly’s test of sphericity was conducted, and since sphericity was violated, the Greenhouse-Geisser correction was applied to adjust the degrees of freedom. Mean and standard deviation were used to summarize knowledge scores at each time point for each intervention group, along with the Wald F statistic and p-value for within-group and between-group comparisons. Post-hoc analyses with Bonferroni correction were performed for significant main effects to identify specific differences between time points. Additionally, a profile plot was generated to visualize changes in knowledge scores over time.

**Procedure to Perform Two-Way Repeated Measures ANOVA in SPSS:**

1. Arrange data in wide format (each row represents one subject with multiple columns for each time point) as given below in Table 1.

Table 1. Data in wide format

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Participant\_ID** | **Group** | **Pre\_test** | **Post\_1** | **Post\_2** |
| 1 | Experimental | 10 | 20 | 26 |
| 2 | Experimental | 14 | 28 | 38 |
| 3 | Experimental | 10 | 27 | 36 |
| 4 | Experimental | 18 | 31 | 36 |
| 5 | Experimental | 19 | 32 | 38 |
| 6 | Experimental | 13 | 24 | 31 |
| 7 | Experimental | 11 | 25 | 35 |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| 41 | Control | 19 | 20 | 21 |
| 42 | Control | 14 | 14 | 16 |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |

1. Go to Analyze > General Linear Model > Repeated Measures
2. Define the within-subject factor (e.g., "Time" with 3 levels)
3. Click "Add" to add the factor to the model
4. Click "Define"
5. Move the outcome variables at different time points to the within-subjects variables box
6. Move the ‘Group’ variable to the between-subjects factor box
7. Click the "Options" button to select descriptive statistics
8. Click the "Plots" button to create interaction plots
9. Click "Continue" and then "OK" to run the analysis

**SPSS Syntax:**

*GLM Pre\_Test Post\_1 Post\_2 BY Group*

*/WSFACTOR=Time 3 Polynomial*

*/METHOD=SSTYPE(3)*

*/PLOT=PROFILE(Time\*Group) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AUTO*

*/EMMEANS=TABLES(OVERALL)*

*/EMMEANS=TABLES(Group) COMPARE ADJ(BONFERRONI)*

*/EMMEANS=TABLES(Time) COMPARE ADJ(BONFERRONI)*

*/EMMEANS=TABLES(Group\*Time) COMPARE(Time) ADJ(BONFERRONI)*

*/PRINT=DESCRIPTIVE ETASQ*

*/CRITERIA=ALPHA(.05)*

*/WSDESIGN=Time*

*/DESIGN=Group.*

**RESULTS:**

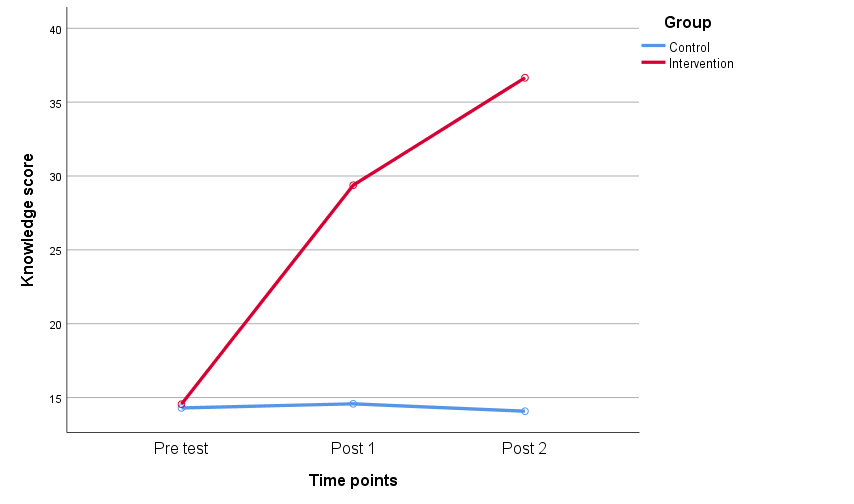
A two-way repeated measures ANOVA was conducted to assess whether there was a significant difference in the average knowledge score across different time points and between the experimental and control groups. The data demonstrated normal distribution, and the assumption of sphericity was not violated.

It is observed that there is a significant increase (p < 0.001) in the average knowledge score in the experimental group, with the mean rising from 14.55 ± 3.04 at Pre-Test to 29.38 ± 4.58 at Post\_1 and further to 36.65 ± 4.45 at Post\_2. In contrast, the control group shows only a modest increase from 14.30 ± 2.85 at Pre-Test to 16.93 ± 3.14 at Post\_1 and 19.17 ± 3.52 at Post\_2, with no significant change observed. This indicates that the intervention had a substantial impact on improving knowledge scores compared to the control.

Also, it was observed that there is a significant interaction between time and groups (p<0.001), that means there is a significant difference in change in knowledge score over time between experimental group and control group. Hence it is concluded that intervention is more effective in increasing knowledge score compared to control group. More details are provided in table2, and profile plot (Fig 1) given below.

**Table 2:** Comparison of Knowledge Scores Between Control and Experimental Groups Across Multiple Time Points

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Groups** | | **Mean** | **SD** | **Within group comparison,**  **F (p value)** | **Between group comparison,**  **F (p value)** |
| **Control group** | Pre\_test | 14.30 | 2.85 | 1.35  (p=0.267) | 879.80 (p<0.001\*) |
| Post\_1 | 16.93 | 3.14 |
| Post\_2 | 19.17 | 3.52 |
| **Experimental**  **group** | Pre\_test | 14.55 | 3.04 | 1253.21  (p<0.001\*) |
| Post\_1 | 29.38 | 4.58 |
| Post\_2 | 36.65 | 4.45 |

****

**Fig.1:** Profile plot- illustrates the interaction between time and experimental group on knowledge scores, clearly demonstrating the differential patterns of change across the three groups over time.

**Post-hoc Analyses (pairwise comparison):**

The Bonferroni pairwise comparison of knowledge scores revealed significant improvements (p<0.001) between all time points (pre-test, post-test 1, and post-test 2) in the experimental group, indicating progressive knowledge acquisition throughout the intervention. In contrast, the control group showed no significant changes in knowledge scores between any time points, demonstrating stable knowledge levels. Details provided below in the Table 3.

Table 3 : Bonferroni pair wise comparison

|  |  |  |
| --- | --- | --- |
| **Pair wise comparison between** | **Bonferroni adjusted p value** | |
| **Experimental group** | **Control group** |
| Pre test & post 1 | <0.001\* | 1.000 |
| Pre test & post 2 | <0.001\* | 1.000 |
| Post 1 & post 2 | <0.001\* | 0.336 |

\*Significant (p<0.05)

**DISCUSSION:**

Two-way repeated measures ANOVA is a widely used statistical technique for analyzing longitudinal data, particularly in clinical research where intervention effects are assessed over time. This method enables researchers to evaluate within-subject and between-group differences while accounting for repeated observations on the same participants10,11. One of its key advantages is increased statistical power, as it reduces within-subject variability by treating each participant as their own control. It also requires fewer participants than between-subject designs while providing more precise estimates. Additionally, RM-ANOVA effectively analyzes both main effects and interactions, allowing researchers to examine how multiple factors influence the outcome. By controlling for individual differences, it enhances the accuracy of detecting true effects, making it particularly valuable for clinical and experimental studies.

Despite its advantages, several methodological challenges must be considered when applying two-way repeated measures ANOVA. A major concern is the violation of sphericity, which can lead to inflated Type I error rates. To address this, Greenhouse-Geisser or Huynh-Feldt corrections should be applied. Alternatively, mixed-effects models offer a more flexible approach by accommodating individual variability in response trajectories and handling missing data more effectively. Missing data, if not properly managed, can introduce bias and reduce statistical power 12. Mixed-effects models or multiple imputation techniques are recommended for handling incomplete data without resorting to listwise deletion.

While two-way repeated measures ANOVA is widely used due to its simplicity and accessibility in statistical software, alternative approaches such as linear mixed-effects models (LMMs) and Bayesian hierarchical models provide greater flexibility. LMMs account for random effects, making them suitable when sphericity assumptions are violated, or subject-specific variability is of interest13. When the normality assumption is violated or covariate adjustment is required, generalized estimating equations (GEE) offer a robust alternative, as they can analyze correlated longitudinal data without strict distributional assumptions14. However repeated measures ANOVA is considered as robust to normality violations, with no impact on Type I error or power, if the sphericity assumption is met15,16.

**Conclusion:**

Two-way repeated measures ANOVA is a robust statistical approach for analyzing longitudinal data in clinical research. Our simulated study on nutritional knowledge levels highlights its capability to evaluate time effects, group differences, and their interaction—key considerations in intervention studies. By adhering to proper assumption testing and interpretation guidelines, researchers can effectively utilize this method to explore complex questions related to changes over time. This paper provides a comprehensive guide to its application, ensuring clinical researchers can confidently implement two-way repeated measures ANOVA in their studies.

**CONFLICT OF INTEREST:**

The authors declare no conflicts of interest related to this research.

**COMPETING INTERESTS DISCLAIMER:**

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

**REFERENCES:**

1. Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. British Journal of Mathematical and Statistical Psychology, 54(1), 1–20.
2. Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1), 13–22.
3. Muhammad, L. N. (2023). Guidelines for repeated measures statistical analysis approaches with basic science research considerations. Journal of Clinical Investigation, 133, e171058.
4. Osborne, J. (2002). Notes on the use of data transformations. Practical Assessment, Research, and Evaluation, 8(1).
5. Pekár, S., & Brabec, M. (2018). Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modeling of correlated data in the behavioral sciences. Ethology, 124(1), 86–93.
6. Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics, 42(1), 121–130.
7. Liu, G., & Liang, K. Y. (1997). Sample size calculations for studies with correlated observations. Biometrics, 53(3), 937–947.
8. Machin, D., Campbell, M. J., Tan, S. B., & Tan, S. H. (2011). Sample size tables for clinical studies (2nd ed.). John Wiley & Sons.
9. Puranik, A., Karun, K. M., & Deepthy, M. S. (2025). Hands-on guide to sample size calculation in medical research using EZR. Journal of Complementary and Alternative Medical Research, 26(3), 119–127.
10. Girden, E. R. (1992). ANOVA: Repeated measures (Vol. 84). Sage.
11. Verma, J. P. (2015). Repeated measures design for empirical researchers. John Wiley & Sons.
12. Enders, C. K. (2010). Applied missing data analysis. Guilford Press.
13. Pinheiro, J. C., & Bates, D. M. (2000). Mixed-effects models in S and S-PLUS. Springer.
14. Karun, K. M., & Deepthy, M. S. (2023). Generalized estimating equations in longitudinal studies: A non-parametric alternative for two-way repeated measures mixed ANOVA. Research Journal of Pharmacy and Technology, 16(9), 2381–2384.
15. Blanca-Mena, M. J., Arnau, J., García-Castro, F. J., Alarcón-Postigo, R., & Bono Cabré, R. (2022). Non-normal Data in Repeated Measures ANOVA: Impact on Type I Error and Power.
16. Kanji, G. K. (1976). Effect of non‐normality on the power in analysis of variance: A simulation study. *International Journal of Mathematical Educational in Science and Technology*, *7*(2), 155-160.