DOI:https://doi.org/10.54809/jkss.vi11.19

گۆڤاری کوردستانیی بۆ لێکۆڵینەوەی ستراتیجیی

Applied Repeated Measures Design in medical field

Kosar Mahmood Hassan

Financial accounting and auditing department, College of Commerce, University of Sulaimani

kosar.hassan@univsul.edu.iq

Abstract

Repeated measures design offers a complete description of the time effect due to the inclusion of all time points. It's used in a wide range of fields such as medical study, behavioral science, agriculture and ecology. A repeated measurement ANOVA, are the appropriate test for making inferences about repeated measures designs, has several advantages. First, it can have greater statistical power, controlling for factors that cause between-subject variability. Second, because of its enhanced statistical power, repeated measures ANOVA can detect the desired effect size with fewer subjects. Finally, it has the ability to track an impact over time, like a task's learning curve. In this paper, the 50 patient's data that had depression and were getting psychotherapy was analyzed using One-way Repeated Measure ANOVA. These data were collected by questionnaire form at three different times: baseline, second month, and fourth month. The results showed a significant difference in mean scores at three-time points, indicating that used treatment had a positive impact on this disease.

Keyword: Repeated measurements, longitudinal data, one way ANOVA, normality test and sphericity test

Recieved: 3/10/2021
Accepted: 27/11/2021

E-ISSN: 2790525-X
P-ISSN: 27905268

1.1 Introduction

Repeated measures analysis deals with repeated measures data, commonly known as longitudinal data, is a type of data in which measurements are recorded on specific subjects repeatedly over a period of time and under different conditions (Singh & Rana and Singhal 2013). Repeated-measures study designs are motivated by a researcher's need or desire to examine the change in an outcome over time. Furthermore, risk factors over time or under different situations may be evaluated repeatedly (Edwards 2000). Repeated-measures study designs are like the completely randomized or randomized blocks designs where subjects are randomized to one of several treatments. The difference is that the response is measured repeatedly during the treatment period instead of once at the end of the treatment period (Lawson 2015). It also includes between-subjects and within-subjects variables, which are commonly utilized in various disciplines such as medicine, psychology, and education. Despite the fact that these designs are typically multivariate, covariance matrix homogeneity and sphericity are satisfied. (Vallejo & Fernandez& Herrero and Conejo 2004)

1.2 Aim of research

The purpose of this paper is to know about the effect of psychotherapy, general term for treating depression by talking about condition and related issues with a mental health professional, on patients who receive this treatment for a period of time.

1.3 Definition of repeated measures and longitudinal data

Repeated measures data is a kind of data that includes a high number of time points as well as changing experimental or observational conditions over the period of the research. Longitudinal data can be regarded as a subset of traditional repeated measures data, with some conceptual differences. They're generally composed of many observations for the same group of people at a limited number of time periods separated by equally or unequally spaced intervals. Therefore, longitudinal data can be described as the results of repeated measurements at a limited number of time points with predetermined time scale, time interval, and other conditions. Longitudinal data can be divided into two types: multivariate and univariate. The data structure for repeated measurements has always followed a multivariate structure (Liu 2015).

1.4 Logic of repeated measures and longitudinal ANOVA

Repeated-measures ANOVA is used to determine if observed differences between treatment conditions are significantly greater than what would be expected if there is no treatment effect. In the numerator of the F-ratio, the between treatments variance measures the real mean differences between the treatment conditions and the variance in the denominator, is also called the error variance because the denominator measures variability caused entirely by random and unsystematic factors, proposed to measure how much difference is sensible to expect if there are no systematic treatment effects and no systematic individual differences (Gravetter 2015).

1.5 Assumption of repeated measures ANOVA table

A research designs are measuring individuals twice or more on the dependent variable. Instead of employing different individuals for each level of treatment, the individuals are given multiple treatments and measured after each one. This indicates that each individual will be their own control. Scores for the same individual are dependent in repeated-measures analysis, but scores for different individuals are independent. The following are the assumptions underlying this type of study design:

1.5.1 Normality- The normality assumption assumes that populations have a normal distribution. Outliers are the most common cause of normality violations. Furthermore, both the ANOVA and MANOVA models are sensitive to outliers. To achieve normality, Data are tested for probable outliers and measured for removal. Graphs (histograms, boxplots), normal probability (Q-Q) plots, skewness, and kurtosis statistics, as well as formal normality tests like as the Shapiro-Wilk test and Kolmogorov–Smirnov test, are used to evaluate univariate normality assumptions. (Telesca 2015).

1.5.2 Variance Sphericity - Sphericity is the condition of variances equality of the differences between all possible pairs of groups. The formal method for detecting violations of the sphericity assumption is Mauchley's test (Kim 2015). The condition of sphericity is not satisfied when this test is significant. However, if this test results in a non-significant result, it is acceptable to assume that the variances of differences are not significantly different (Quinn & Keough 2002). In the case of unsatisfied sphericity condition, a correction, such as Greenhouse-Geisser, Huynh-Feldt, and Lower Bound, can be achieved by correcting the degrees of freedom. These corrections estimate epsilon and multiply the numerator and denominator degrees of freedom by this estimate before determining the significance levels for the F tests (Field 2000).

1.5.3 Independence- It means independency of the error and the observations, each experimental unit is independent of each other experimental unit,. The Violation of this assumption can lead to increased Type I error rate. The group's interaction may impact the members' scores, resulting in correlated observations. Correlated observations lead to an overestimation of true probability, which can be corrected by testing at a higher level of significance (Nimon & Williams 2009).

1.6 Hypothesis for repeated measures

The null hypothesis will test that there is no difference between the various groups i.e. all treatment effects are zero. It can state that null hypothesis as;

H_0: μ_1= μ_2=...= μ_p

The alternative hypothesis states that there are significant differences between the treatments conditions, the treatments do have different effects, which means it may be responsible for causing mean differences between the samples. The alternative hypothesis can show as follows;

If the p-value represents the probability of getting a sample outcome, given that the value stated in the null hypothesis is true. When the p-value is less than 5% the null hypothesis is rejected and it accept when the p value is greater than 5% (Gravetter 2015).

1.7 Analysis of variance (ANOVA) table

Analysis of variance (ANOVA) is a statistical technique to analyze variation in a response variable (continuous random variable) measured under conditions defined by discrete factors (classification variables, often with nominal levels). Frequently, we use ANOVA to test equality among several means by comparing variance among groups relative to variance within groups (random error). Sir Ronald Fisher pioneered the development of ANOVA for analyzing results of agricultural experiments. Today, ANOVA is included in almost every statistical package, which makes it accessible to investigators in all experimental sciences. It is easy to input a data set and run a simple ANOVA, but it is challenging to choose the appropriate ANOVA for different experimental designs, to examine whether data adhere to the modeling assumptions, and to interpret the results correctly (Larson 2015).

1.8 One way repeated measures ANOVA table

The simplest design between repeated measure designs is one way repeated measure design. This design has one factor, and all of the experimental units are tested within the factor levels. Various time points (periods), varying treatments, or different level of the same treatment are examples of repeated factors. The design's model is,

y_ij= μ+ a_i+ β_j+ ε_(ij)   i=1,2,3,..,n   ;j=1,2,…,k

Where μ is the overall mean, αi represents the impact of an i^{th} subject, βj represents the effect of the j^{th} time or treatment and ε_{ij} is an error term.

The sources of total variation in one-way Repeated Measure Design are the conditions (times) and within-group variations, and the goal is to test the differences between the periods or treatments. Variance among subjects and error are other sources of within-subject variation. Thus,

The following formula is used to compute all sources of variation;

A repeated measure ANOVA similarly calculates an F-statistic:

$$SS_{Total} = SS_{Conditions} + SS_{Within-Subject}$$

$$= SS_{Conditions} + SS_{Subjects} + SS_{Error} \ldots (١)$$

$$SS_{Error} = SS_{Within-Subject} - SS_{Subjects} \quad \ldots (٢)$$

The following formula is used to compute all sources of variation;

$$SS_{Conditions} = \sum_{i=١}^{k} n_i (\bar{x}_i - \bar{x})^٢ \quad \ldots (٣)$$

$$SS_{Within} = \sum_{١} (x_{i1} - \bar{x}_١)^٢ + \sum_{٢} (x_{i2} - \bar{x}_٢)^٢ + \cdots + \sum_{k} (x_{ik} - \bar{x}_k)^٢ \quad \ldots (٤)$$

$$SS_{Subjects} = k . \sum_{i=١}^{k} (\bar{x}_i - \bar{x})^٢ \quad \ldots (٥)$$

Where;

$$SS_{Total} = \sum_{i=١}^{n} \sum_{j=١}^{k} x_{ij}^٢ - \frac{\left(\sum_{i=١}^{n} \sum_{j=١}^{k} x_{ij}\right)^٢}{nk} \quad \ldots (٦)$$

Table 1 presents a general illustration of one-way repeated measures designs with n subject and p treatments or periods (repeated measures).

Table 1- Tabular Presentation of a Repeated Measures ANOVA

$$F = \frac{MS_{conditions}}{MS_{error}} \quad \ldots (٦)$$

$$MS_{conditions} = \frac{SS_{conditions}}{k - ١} \quad \ldots (٧)$$

$$MS_{error} = \frac{SS_{error}}{(k - ١)(n - ١)} \quad \ldots (٨)$$

Source	SS	df	MS	F
Conditions	$SS_{conditions}$	(k-1)	$MS_{conditions}$	
Subject	$SS_{subject}$	(n-1)	$MS_{subject}$	
Error	SS_{error}	(k-1)(n-1)	MS_{error}	
Total	SS_{Total}	(N-1)		

Another measure in the ANOVA model is partial eta squared, which is used to calculate the size of the effect of various factors. It measures the proportion of variance explained by a specific variable that remains after accounting for variance explained by other variables in the model. The following is the formula for calculating partial eta squared: (Görgülü & Mokhles and Şahinler 2006)

$$\text{Partial eta squared } (\eta^٢) = \frac{SS_{conditions}}{SS_{conditions} + SS_{error}} \quad \ldots (٩)$$

Data description and analysis

2.1 Data description

Depression is a mood disorder that affects a person's capacity to perform at home or at work due to overwhelming feelings of sadness, loss, and hopelessness. People with Parkinson's disease have abnormalities in certain neurotransmitters that regulate mood and are thought to play a significant role. The common factors of depression are mental, biological, and environmental. One of depression treatment method is psychotherapy and there are some ways to Diagnosis of depression score; physical examination, laboratory tests, DSM-5 and psychological evaluation. This paper aimed to know about the effect of psychotherapy on patients. Psychological evaluation was used to collect the data of this paper by filling out a questionnaire form which contained some question about Symptoms, thoughts, feelings and behavior patterns of patients. This form was filled by 50 patients who had depression and received this treatment for four months. The results were as shown in appendix A.

Descriptive statistics for the depression scores of patients at three different time point presents in Table 2.

Table 2: Descriptive statistics for depression scores

		n	Std. deviation
		0	9.579
Second month	50	19.52	10.545
Forth month	50	16.74	11.611

By depending on table 2, the mean scores of depression decreased i.e. the treatment had a positive impact on patients.

2.2 Analysis of One way Repeated Measure ANOVA

2.2.1 Normality test

When using repeated-measures ANOVA to compare three or more observations of a continuous outcome, the assumption of normality of difference scores must be checked. In each cell in the design, the outcome variable should be approximately normally distributed. The Kolmogorov-Smirnova Shapiro-Wilk normality test can be used to verify this.

Table 3 - Tests of Normality

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	.Sig	Statistic	df	.Sig
Baseline	0.102	50	0.200	0.970	50	0.228
Second Month	0.102	50	0.200	0.966	50	0.157
Fourth Month	0.083	50	0.200	0.948	50	0.029

Table 3 displays the results from the two used tests for each measurement separately. The Shapiro-Wilk Test is more appropriate for small sample sizes (< 50 samples), but can also handle large sample sizes. These results demonstrated that the assumptions were satisfied for each .measurement and the used data is normally distributed because $p > 0.01$

To determine normality graphically, we can use the output of a normal Q-Q plot and histo-.gram. As shown below, the data was also graphically distributed normal distribution

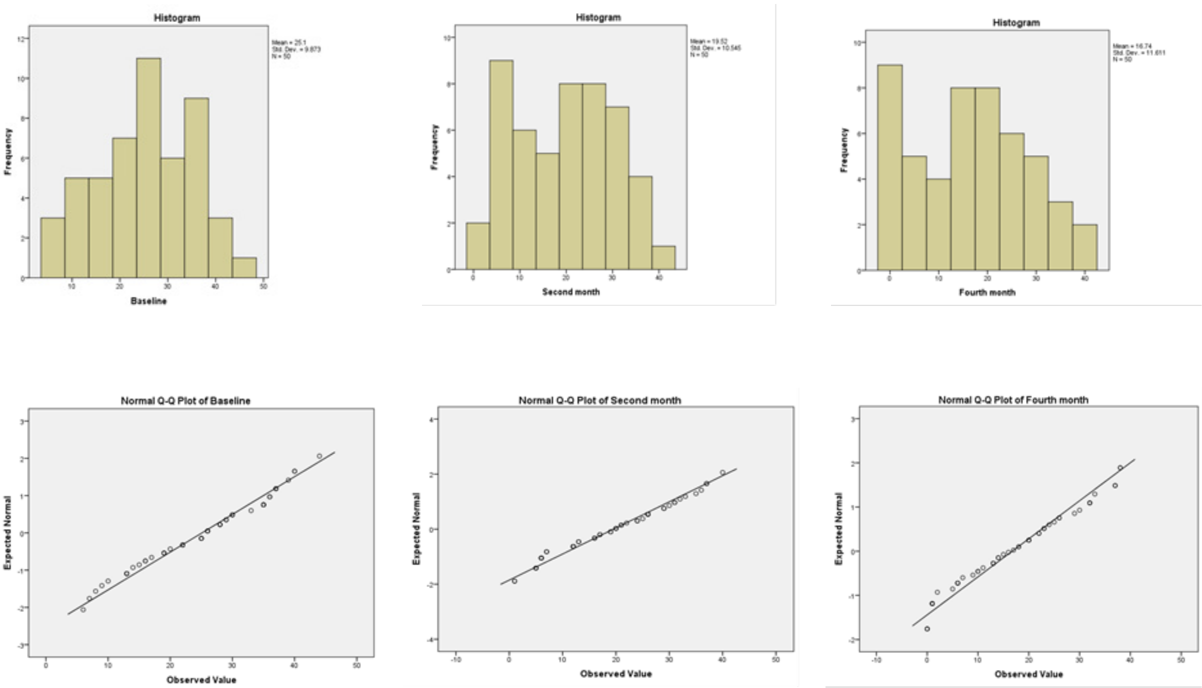

Figure 1- Histograms and Q-Q plots of three different time

Sphericity Test 2.2.2

To test the Sphericity (the variances of the differences are equal), Mauchly’s Test was used. ,Results from this Test of are shown below

Table 4: Sphericity test result

Within subjects effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	0.944	2.790	2	0.248	0. 947	0.983	0. 500

By depend on table 4 results, sphericity has not been met ($p = 0.022$). As a result, the null hypothesis will be accepted. To solve this problem, corrections to the degrees of freedom must be performed so that a valid critical F-value can be obtained. As mentioned, three corrections are used;

Greenhouse-Geisser Correction- As illustrated below, the Geisser technique calculates epsilon to adjust the degrees of freedom of the F-distribution:

df_conditional= ε □(k-1)

df_conditional= 0.947 (3-1)=1.89

df_error= ε □(k-1)(n-1)

df_error= 0.947 (3-1)(50-1)= 92.8

A correction has increased significance value accuracy. It has also adjusted the p-value to compensate for the fact that when sphericity is violated, the test is too large.

Huynd- Feldt Correction- The Huynd- Feldt correction, like the Greenhouse-Geisser correction, calculates epsilon (represented as $\tilde{\varepsilon}$) to adjust F-distribution's degrees of freedom , as illustrated below:

This correction also has caused a more accurate significance value, and the same is true for the Lower bound correction. These corrections have no effect on the F-statistic because the estimated epsilon is multiplied by the degrees of freedom in both the numerator and denominator, canceling each other out, as demonstrated below:

After all these corrections the one-way ANOV become as follows:

Table 5: Test of within-subject effects

Source		Type III sum of squares	df	Mean square	F value	P value	Partial Eta squared
Conditions (Time)	Sphericity Assumed	1812.573	2	906.287	27.656	0.000	0.36
	Greenhouse-Geisser	1812.573	1.893	957.458	27.656	0.000	
	Huynh-Feldt	1812.573	1.967	921.528	27.656	0.000	
	Lower-bound	1812.573	1.000	1812.573	27.656	0.000	
Error	Sphericity Assumed	3211.427	98	32.770			
	Greenhouse-Geisser	3211.427	92.762	34.620			
	Huynh-Feldt	3211.427	96.379	33.321			
	Lower-bound	3211.427	49.000	65.539			

The actual results of the repeated measures ANOVA are presented in the above table (Tests of Within-Subjects Effects table). It demonstrates that there was a statistically significant difference between the means at the various time periods. This table displays the F-value for the time factor, as well as the significance level and impact size (Partial Eta squared) associated with it.

After adjusting the degrees of freedom for the sphericity assumed condition and depending on the results of the within-subjects ANOVA, we reject the null hypothesis. Thus, there was a significant difference in the depression score overtime with a relatively large effect size (partial η^2 =0.36).

3. Conclusion

Repeated measures ANOVA provides researchers with ways to test research hypotheses by controlling for subject variance. It is also important in light of several advantages: statistical power, fewer subjects, and efficiency. As one of the parametric analyses, Repeated Measures ANOVA can be powerful because it can control for factors that cause between subjects variability. It can identify the appropriate effect size with fewer individuals. Since fewer subjects may be sufficient for the analysis, it is quicker and cheaper to recruit, train, and compensate to complete an entire experiment. A one-way repeated-measures ANOVA was used in this study to see if there were statistically significant differences in depression scores over the course of 4-month psychotherapy. Results showed that the data were normally distributed for each time, as measured by numerical method (Shapiro-Wilk and Kolmogorov-Smirnova test) and graphical method (boxplot and histogram) respectively. The assumption of sphericity was assessed by Mauchly's Test of Sphericity, p = 0.248, and was not violated. Therefore, a three method of correction (Greenhouse-Geisser, Huynh-Feldt and Lower-bound) was applied (Ɛ = 0.947, 0.983 and 0.500). The psychological therapy caused statistically significant changes in depression score over time with decreasing from baseline (M = 24.90, SD = 9.579) to 2 months (M = 19.52, SD = 10.545) and 4 months (M = 16.74, SD = 11.611).

پوختە

نەخشەکێشانی پێوانە دوبارەییەکان وەسفێکی تەواوەتی پێشکەش دەکات بۆ کاریگەری کات بە ڕەچاو کردنی زیادکردنی هەمو خاڵە کاتییەکان. کە بەکاردێت لە بوارە جۆراوجۆرەکانی وەک توێژینەوەی پزیشکی، زانستە ڕەفتارییەکان، کشتوکاڵ و ژینگە. نەخشەکێشانی پێوانەی (ANOVA) تاقیکاریەکی گونجاوە بۆ گەیشتن بە دەرەنجام دەربارەی پێوانە دوبارەییەکان. کۆمەڵێک سود و تایبەتمەندی هەیە لەوانە توانایەکی گەورەی ئاماری، دەست بەسەراگرتنی هۆکارەکانی جیاوازی نێوان بابەتەکان. بەهۆی

توانـای ئامـاری بـاڵای، خشـتەی ANOVA بـۆ پێوانـە دوبارەییـەکان توانـای دیـاری کردنـی ڕێـژەی گۆڕانـی پێویسـتی هەیـە لەگـەڵ چەنـد بابەتێکی کـەم. لـە کۆتایـدا، توانـای بەدواداچونـی گۆڕانـکاری بـە تێپەڕبونـی کات، هاوشـیوەی چەمـاوەی فێربـون. لـەم توێژینەوەیـەدا (One way ANOVA) بـۆ پێوانـە دوبارەییـەکان بەکارهاتـوە بـۆ شـیکاری زانیـاری پەنجـا توشـبوی نەخۆشـی خەمۆکـی کـە چارەسـەری دەرونیـان وەرگرتـوە . زانیاریـەکان لـە سـێ کاتـی جیـاوازدا و بـە بەکارهێنانـی فۆڕمـی ڕاپرسـی کۆکراوەتـەوە: پێـش چارەسـەر، مانگـی دووەم و چـوارەم لـە کاتـی وەرگرتنـی چارەسـەر. وە ئەنجامـەکان دەریـان خسـتوە کـە جیـاوازی هەیـە لـە نێـوان تێکـرای پلـەی خەمۆکـی لـەو سـێ کاتـەدا، ئەمـەش ئـەوە دەگەیەنێـت کـە ئـەم چارەسـەرە کاریگـەری ئەرێنـی هەبـوە بـۆ چارەسـەری ئـەم نەخۆشـیە.

الملخص

يقـدم تصميـم المقاييـس المتكـررة وصفًـا كامـلاً لتأثـير الوقـت نظـرًا لإدراج جميـع النقـاط الزمنيـة. يتـم اسـتخدامه في المجـالات المتنوعـة مثـل البحـث الطبـي ، العلـوم السـلوكية ، الزراعـة والبيئـة. تحليـل التبايـن للقياسـات المكـررة ، هـو الاختبـار المناسـب للتوصـل إلى اسـتنتاجات حـول تصميـمات المقاييـس المتكـررة ، ولـه عـدد مـن المزايـا. أولاً ، يمكـن أن يكـون لهـا قـوة إحصائيـة أكـبر ، والتحكـم في العوامـل التـي تسـبب بـين تقلـب الموضـوع. ثانيًـا ، نظـرًا لقوتهـا الإحصائيـة المعـززة ، يمكـن لجـدول تحليـل التبايـن للقياسـات المكـررة اكتشـاف حجـم التأثـير المطلـوب مـع عـدد أقـل مـن الموضوعـات. أخـيرًا ، يمكنـه تتبـع تأثـير بمـرور الوقـت ، مثـل منحنـى تعلـم المهمـة. في هـذه الورقـة ، تـم تحليـل البيانـات ٥٠ مريـض واللذيـن يعانـون مـن الاكتئـاب وكانـوا يتلقـون العـلاج النفسـي باسـتخدام طريقـة تحليـل التبايـن الاحـادى . و تـم جمـع هـذه البيانـات عـن طريـق اسـتمارة الاسـتبيان في ثـلاث أوقـات مختلفـة: قبـل العـلاج ، الشـهر الثـاني والشـهر الرابـع. أظهـرت النتائـج فرقـاً معنويـاً في متوسـط الدرجـات في ثـلاث نقـاط زمنيـة ، مـما يشـير إلى أن هـذا العـلاج كان لـه تأثـير إيجـابي عـلى هـذا المـرض.

4. References

1. Ahmadi, H. (2019), A Three-Paper Dissertation on Longitudinal Data Analysis in Education and Psychology, PhD thesis, Columbia University.
2. Edwards, L. (2000) 'Modern Statistical Techniques for the Analysis of Longitudinal Data in Biomedical Research', Pediatric Pulmonology journal, 30(4), pp. 330- 344.
3. Field, A. (2000) 'A Bluffer's Guide to ... Sphericity Andy Field University of Sussex', BPS-MSC Newsletter journal, 6 (1), pp. 13.
4. Görgülü, Ö. Mokhles, Şahinler, S. (2006) 'Repeated Measures Analysis and Some Experimental Design Considerations In Animal Science', ResearchGate journal, pp. 76- 97.
5. Gravetter, F. Wallnau, L. 2015. Introductory Statistics for The Behavioral Sciences. 10th ed. Cengage Learning, nelson Education, Ltd, Canada.
6. Hickey, G. Mokhles, M. Chambers, C. Kolamunnage-Dona R. (٢٠١٨) 'Statistics primer: Performing repeated – measures analysis', Interactive CardioVascular and Thoracic Surgery journal, 26 (4), pp. 539- 544.
7. Kim, H. (2015) 'Statistical notes for clinical researchers: A one- way repeated measures ANOVA for data with repeated

observations', Restorative Dentistry & Endodontics journal, 40(1), pp. 91- 95.
8. Larson, G. (2015) 'Analysis of Variance'. Circulation journal, 117(1), pp. 115- 121.
9. Lawson, J. 2015. Design and Analysis of Experiments with R. CRC Press, Taylor & Francis Group, LLC, Canada.
10. Liu, X. (2015) 'Introduction to longitudinal data analysis in psychiatric research'. Shanghai Archives of Psychiatry Journal, 27(4), pp. 256-259.
11. Nimon, K. Williams, C. (2009) 'Evaluating Performance Improvement through Repeated Measures: A Primer for Educators Considering Univariate and Multivariate Designs'. Research in Higher Education Journal, 16(3), pp. 1-20.
12. Park, E. (2020), 'The Comparison Between Repeated Measures Anova And Multilevel Modeling To Investigate The Effectiveness Of Language Instruction', Advances in Social Sciences Research Journal, 7(6), pp. 426-437.
13. Quinn, G. Keough, M. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, New York.
14. Singh, V. Rana, R. Singhal, A. (2013) 'Analysis of repeated measurement data in the clinical trials', journal of Ayurveda and integrative medicine, 4(2), pp. 77-81.
15. Telesca, L. 2015. 'The Effect of a Metalinguistic Approach to Sentence Combining on Written Expression in Eighth Grade Science for Students who Struggle with Literacy'. PhD thesis, University of Central Florida, Orlando, Florida.
16. Vallejo, G. Fern·ndez, P. Herrero, F. Conejo, N. (2004) 'Alternative procedures for testing fixed effects in repeated measures designs when assumptions are violated'. Psicothema Journal, 16(3), pp. 498-508.

5. Appendix

Appendix A- Data description

Person No.	Baseline	Second month	Fourth month
1	19	19	20
2	13	12	13
3	29	31	23
4	25	26	22
5	10	7	6
6	28	21	9
7	33	16	7
8	25	26	22
9	26	6	14
10	35	31	38
11	36	29	32
12	39	35	32
13	22	22	14
14	37	26	18
15	6	1	1
16	28	1	11
17	25	20	20
18	16	6	2
19	19	19	18

20	30	32	32
21	25	26	23
22	29	25	26
23	15	12	1
24	36	29	30
25	26	17	20
26	28	21	10
27	35	5	0
28	26	6	15
29	37	37	37
30	20	13	1
31	30	13	6
32	35	24	17
33	17	16	16
34	14	26	0
35	37	33	38
36	35	30	6
37	30	37	37
38	22	20	25
39	13	12	13
40	40	36	33
41	13	5	5
42	16	16	10
43	34	24	29
44	9	7	1
45	40	40	26
46	22	20	24
47	7	6	1
48	8	5	0
48	19	12	13
50	26	17	20