drtamil@gmail.com

Repeated Measurement Analysis

GLM Repeated Measures

drtamil@gmail.com

Introduction

- The simplest repeated measurement analysis is the pre-post type of study, where we have only two timepoints.
- There are many situations where one collects information at baseline and then at regular intervals over time, say three monthly, and is interested to determine whether a treatment is effective over time.

drtamil@gmail.com

Common techniques

1. Mean response over time – Interest in overall treatment effect. No information on treatment effect changes over time.
2. Separate analyses at each time point – This is most common in medical journals. Repeated testing at each time point causes inflated type I error and results in interpretation problems. Treatment standard errors are less accurate as only observations at each time point used. Must be discouraged!
3. Analyses of response features – Area under the curve, minimum/maximum values, time to max values.

drtamil@gmail.com

Let us consider a dataset from SPSS (Table I) where the number of errors made by each subject as each repeats the same task over 4 trials were recorded.

Table I. Anxiety data set (Longitudinal form).

Subject	Anxiety	Trial 1	Trial 2	Trial 3	Trial 4
1	Low	18	14	12	6
2	Low	19	12	8	4
3	Low	14	10	6	2
4	Low	16	12	10	4
5	Low	12	8	6	2
6	Low	18	10	5	1
7	High	16	10	8	4
8	High	18	8	4	1
9	High	16	12	6	2
10	High	19	16	10	8
11	High	16	14	10	9
12	High	16	12	8	8

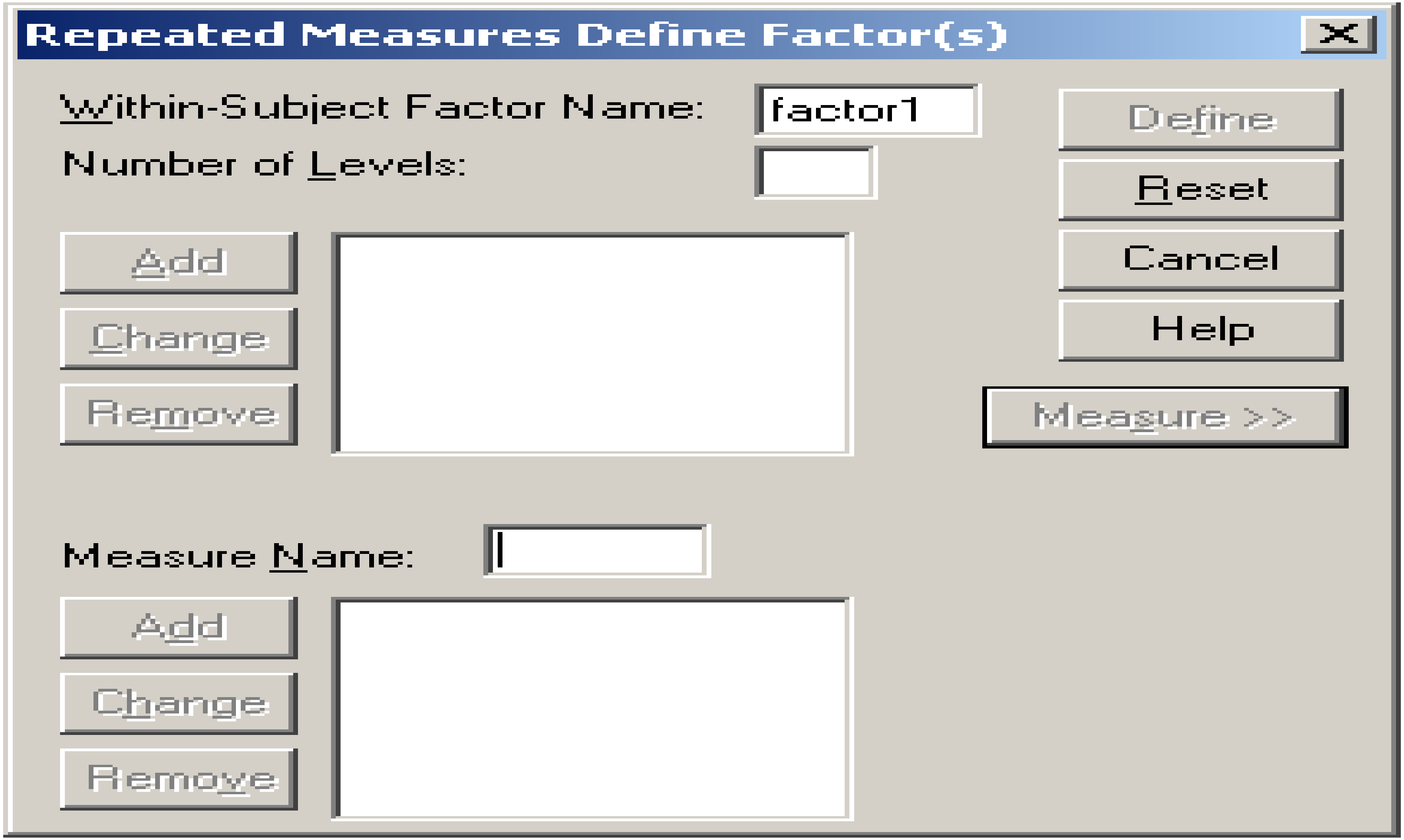
drtamil@gmail.com

Three questions one would want to ask are:

1. Is there a difference in the number of errors made between the Low and High anxiety subjects? This is termed as the Between-Subject Factor – a factor that divides the sample of subjects into distinct subgroups.
2. Is there a reduction in the number of errors made over trials – a time trend? This is termed as the Within-Subject Factor - distinct measurements made on the same subject, for example, BP over time, thickness of the vertebrae of animals.
3. Is there a group time interaction? If there is a time trend, whether this trend exists for all groups or only for certain groups?

drtamil@gmail.com

Analyse, General Linear Model, Repeated Measures



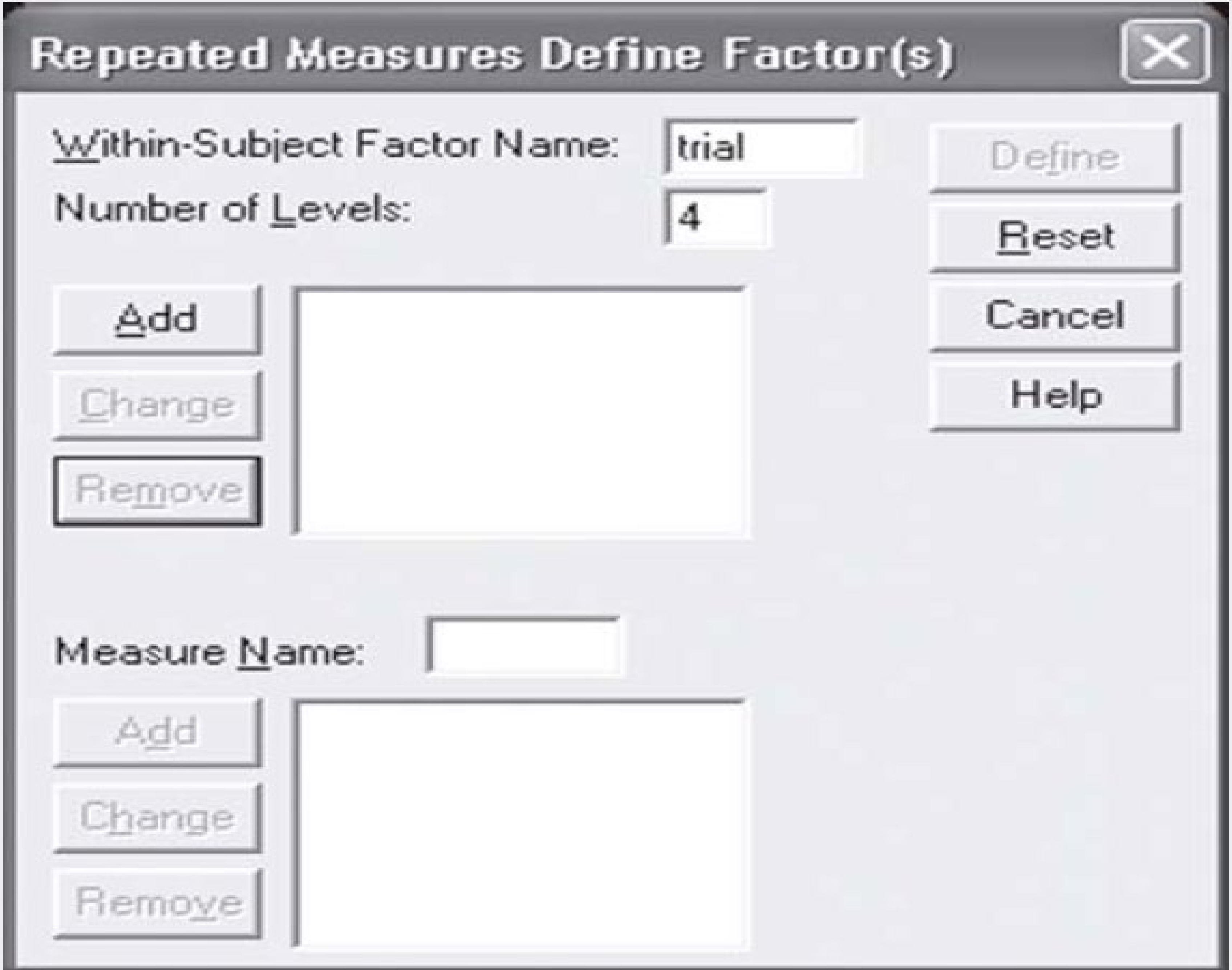
drtamil@gmail.com

GLM – Repeated Measures

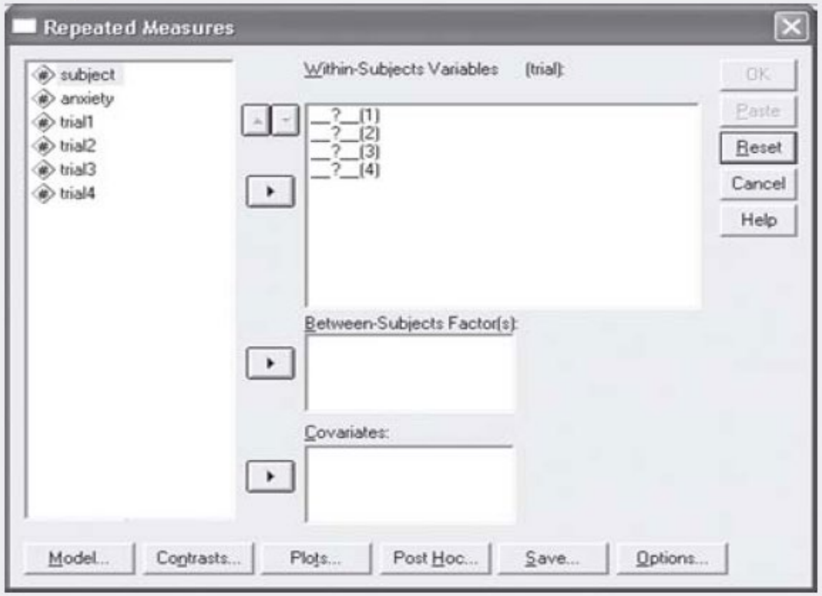
- Change the Within-Subject Factor Name to “trial” (or any suitable term) and put “4” in the Number of Levels (number of repeated measurements) – see Template II.

drta

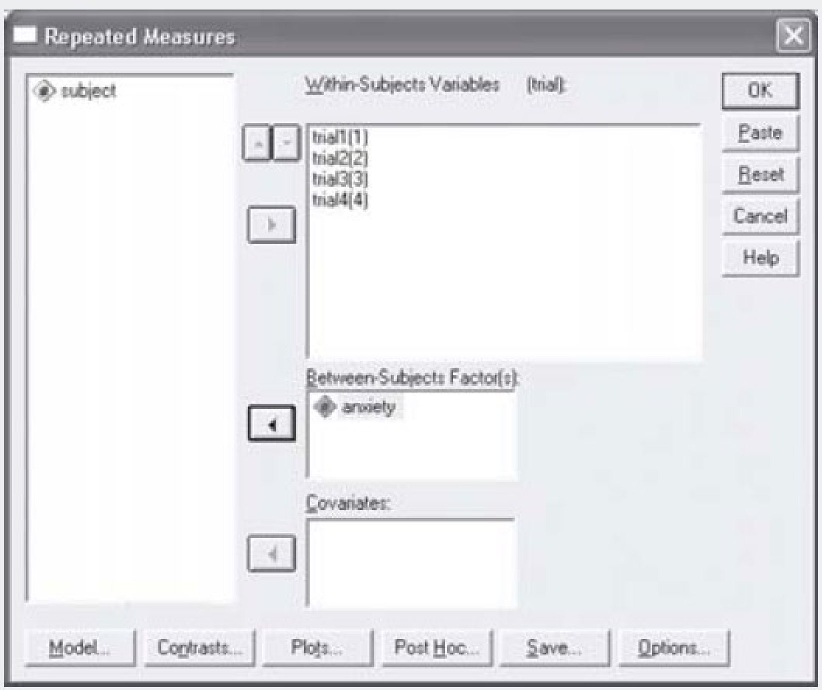
Template II. Defining the number of levels.



Template III.



Template IV.



1. Click Add, then Define.
2. Bring the variables “trial1” to “trial4” over to within-Subjects Variables panel and “anxiety” to the Between-Subjects Factor panel.
3. The above steps set up the “basic” analyses for a repeated measurement analysis.

drtamil@gmail.com

Results

1. THE BETWEEN-SUBJECTS DIFFERENCE

Table IIa. Between-Subjects difference.

Tests of Between-Subjects effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III sum of squares	df	Mean square	F	Sig.
Intercept	4800.000	1	4800.000	280.839	.000
Anxiety	10.083	1	10.083	.590	.460
Error	170.917	10	17.092		

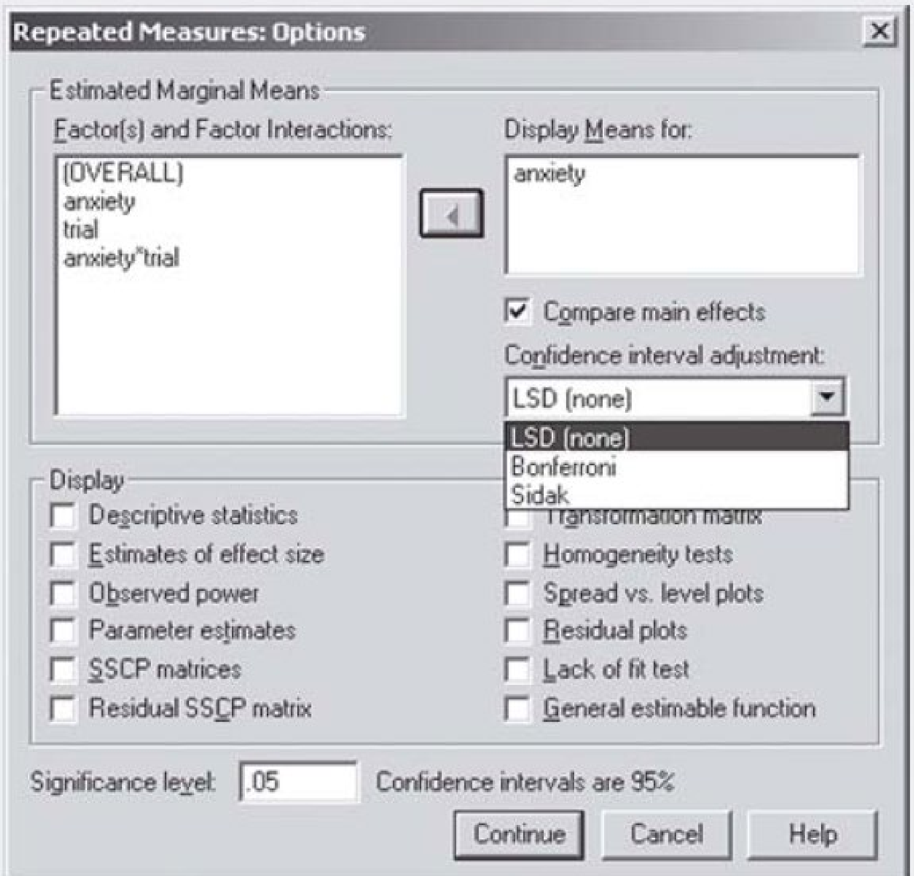
Table IIa shows that there were no differences in the mean number of errors made over time between the Low and High anxiety groups (p=0.460).

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.521	1	2.521	.590	.460
Within Groups	42.729	10	4.273		
Total	45.250	11			

drtamil@gmail.com

Further Analysis

Template V. Options for Comparing Main effects.



- Put "anxiety" in the Display Means panel- this will give Table IIb. To get Table IIc, tick the Compare main effects box and choose Bonferroni (using the most conservative technique to adjust the p value for multiple comparisons(4)).
- The LSD (none) does not adjust the p value for the multiple comparisons.

drtamil@gmail.com

Results

Table IIb. Descriptive statistics by anxiety.

Anxiety				
Measure: MEASURE_1				
			95% Confidence interval	
Anxiety	Mean	Std. error	Lower bound	Upper bound
Low anxiety	9.542	.844	7.661	11.422
High anxiety	10.458	.844	8.578	12.339

drtamil@gmail.com

Results

Table IIc. Pairwise comparisons by anxiety.

Pairwise Comparisons

Measure: MEASURE_1

(I) Anxiety	(J) Anxiety	Mean difference (I-J)	Std. error	Sig.[a]	95% Confidence interval for Difference[a] Lower bound	Upper bound
Low anxiety	High anxiety	-.917	1.193	.460	-3.576	1.742
High anxiety	Low anxiety	.917	1.193	.460	-1.742	3.576

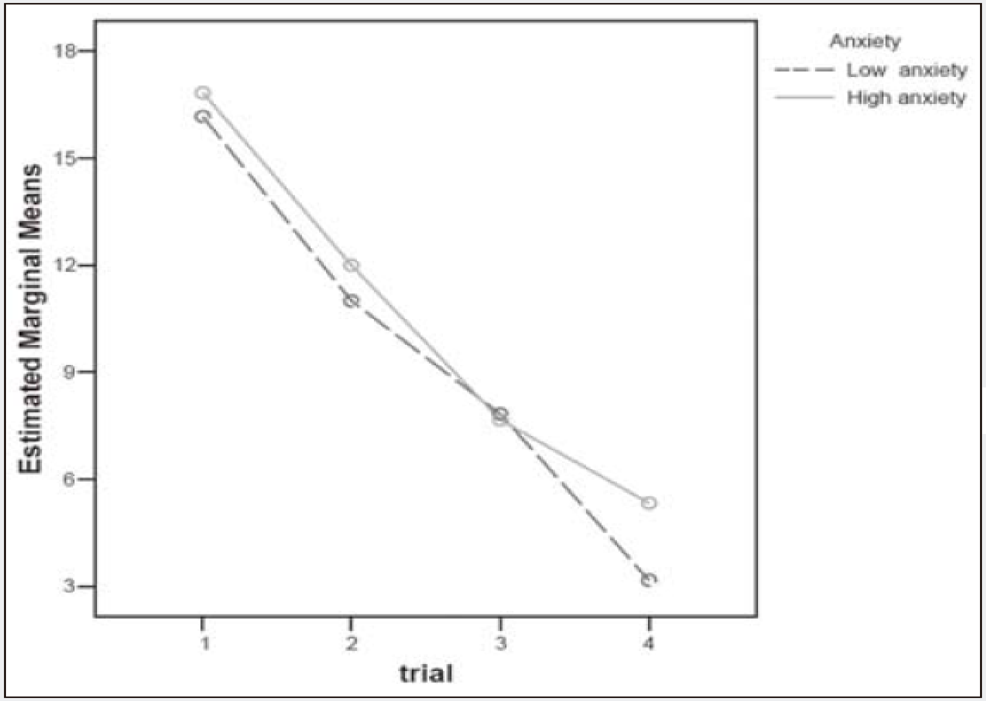
Based on estimated marginal means.

[a] Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

drtamil@gmail.com

Plots

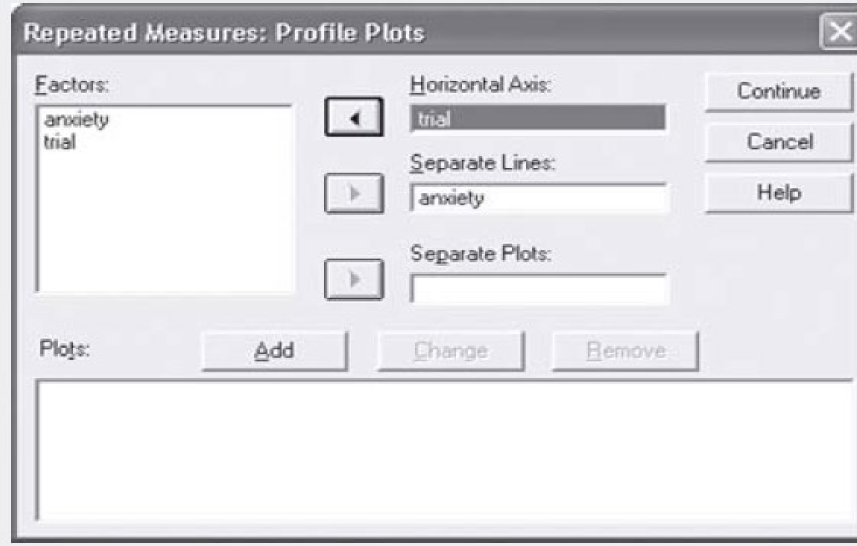
Fig. 1. Graphical plot for repeated measurement analysis



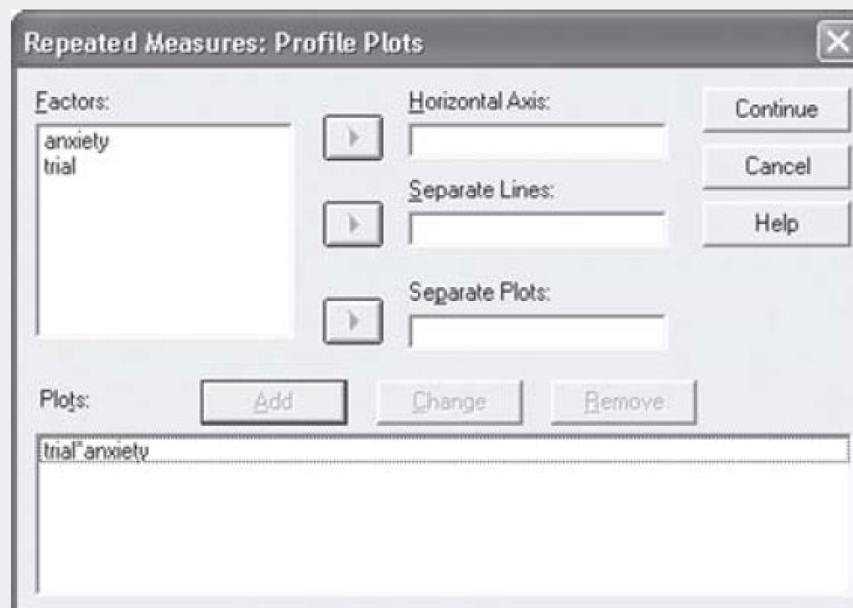
To get a helpful graphical plot (Fig. 1), click on the Plots folder in Template IV to get Template VII.

Put "trial" in the Horizontal Axis and "anxiety" in the Separate Lines – the Add button becomes visible, click on it to get Template VIII.

Click Continue and then click on OK in Template IV to run the analysis.



Template VIII. Requesting for plots.



drtamil@gmail.com

Within Subjects Analysis

Table IIIa. Descriptive statistics of trial by anxiety.

	Descriptive statistics			
	Anxiety	Mean	Std. deviation	N
Trial 1	Low anxiety	16.17	2.714	6
	High anxiety	16.83	1.329	6
	Total	16.50	2.067	12
Trial 2	Low anxiety	11.00	2.098	6
	High anxiety	12.00	2.828	6
	Total	11.50	2.431	12
Trial 3	Low anxiety	7.83	2.714	6
	High anxiety	7.67	2.338	6
	Total	7.75	2.417	12
Trial 4	Low anxiety	3.17	1.835	6
	High anxiety	5.33	3.445	6
	Total	4.25	2.864	12

- Both anxiety groups do display a reduction in the number of errors over time, as observed from Fig. 1.
- Is this reduction trend significant for both groups or just for one group?
- Repeated measurement analysis give us 2 “approaches” to analyse the Within-Subjects effect: **Univariate** and **Multivariate** (both approaches give the same result for the Between-Subject effect).

drtamil@gmail.com

Univariate Approach

GLM

Univariate

- The **Univariate** approach needs the Within-Subjects variance-covariance to have a Type H structure (or circular in form – correlation between any two levels of Within-Subjects factor has the same constant value). This assumption is checked using the Mauchly's Sphericity test (Table IIIb).

drtamil@gmail.com

Mauchly's test of Sphericity

Table IIIb. Sphericity test.

Mauchly's test of Sphericity[b]

Measure: MEASURE_1

					Epsilon[a]		
Within-Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Trial	.283	11.011	5	.053	.544	.701	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalised transformed dependent variables is proportional to an identity matrix.

[a] May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

[b] Design: Intercept + anxiety
Within Subjects Design: trial

We want the Sig to be >0.05 for the assumption of sphericity to be valid. If Sig <0.05, we can use the adjusted p values given by Greenhouse-Geisser, Huynh-Feldt or Lower-bound.

drtamil@gmail.com

Table IIIc. Univariate test of Within-Subjects effects.

Tests of Within-Subjects effects

Measure: MEASURE_1

Source		Type III sum of squares	df	Mean square	F	Sig.
Trial	Sphericity Assumed	991.500	3	330.500	128.627	.000
	Greenhouse-Geisser	991.500	1.632	607.468	128.627	.000
	Huynh-Feldt	991.500	2.102	471.773	128.627	.000
	Lower-bound	991.500	1.000	991.500	128.627	.000
Trial * anxiety	Sphericity Assumed	8.417	3	2.806	1.092	.368
	Greenhouse-Geisser	8.417	1.632	5.157	1.092	.346
	Huynh-Feldt	8.417	2.102	4.005	1.092	.357
	Lower-bound	8.417	1.000	8.417	1.092	.321
Error (trial)	Sphericity Assumed	77.083	30	2.569		
	Greenhouse-Geisser	77.083	16.322	4.723		
	Huynh-Feldt	77.083	21.016	3.668		
	Lower-bound	77.083	10.000	7.708		

Table IIIc shows that there is a reduction of errors committed over trials ($p<0.001$ given by the Sig value of the Source = trial with sphericity assumed).

The Sig of source = trial*anxiety with sphericity assumed is 0.368 which means that there is no time*group interaction, i.e. both low and high anxiety groups had a reduction in the number of errors made over trials.

drtamil@gmail.com

Multivariate Approach

GLM

drtamil@gmail.com

Multivariate Approach

- The **Multivariate** approach assumes that the correlation for each level of Within-Subjects factor is different and the vector of the dependent variables follows a multivariate normal distribution with the variance-covariance matrices being equal across the cells formed by the Between-subject effects.
- This homogeneity of the Between-Subjects variance-covariance is checked by using Box's M test (Table IIId); obtained by ticking the Homogeneity test box in Template V.

drtamil@gmail.com

Homogeneity Test

Table IIId. Box's M test.

Box's test of equality of Covariance Matrices[a]	
Box's M	21.146
F	1.161
df1	10
df2	478.088
Sig.	.315

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

[a] Design: Intercept + anxiety
Within-Subjects design: trial

- The p value for the Box's test is 0.315 (we want $p>0.05$), implying that the homogeneity assumption holds.

drtamil@gmail.com

Table IIIe. Multivariate test of Within-Subjects effects.

Multivariate tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
Trial	Pillai's Trace	.961	64.854[a]	3.000	8.000	.000
	Wilk's Lambda	.039	64.854[a]	3.000	8.000	.000
	Hotelling's Trace	24.320	64.854[a]	3.000	8.000	.000
	Roy's Largest Root	24.320	64.854[a]	3.000	8.000	.000
Trial * anxiety	Pillai's Trace	.479	2.451[a]	3.000	8.000	.138
	Wilk's Lambda	.521	2.451[a]	3.000	8.000	.138
	Hotelling's Trace	.919	2.451[a]	3.000	8.000	.138
	Roy's Largest Root	.919	2.451[a]	3.000	8.000	.138

[a] Exact statistic
[b] Design: Intercept + anxiety
Within-Subjects design: trial

- Table IIIe shows the Within-Subjects analysis from the Multivariate procedure. Once again, there is a time trend effect ($p<0.001$) with no time*group interaction effects ($p=0.138$).
- Most of the time the results from Pillai's Trace, Wilks' Lambda, Hotelling's Trace and Roy's Largest Root should be the similar. In the event when the results are different, Wilks' Lambda should be chosen.

drtamil@gmail.com

What’s the difference?

- The multivariate tests table displays four tests of significance for each model effect. In analogy to univariate tests, the "ratio" of the hypothesis SSCP matrix to the error matrix is used to evaluate the effect of interest. More specifically, the eigenvalues of the test matrix defined by the matrix product of the appropriate hypothesis SSCP matrix and the inverse of the error SSCP matrix are used to compute the statistics in the multivariate tests table.

drtamil@gmail.com

How were they derived?

- Pillai's trace is a positive-valued statistic. Increasing values of the statistic indicate effects that contribute more to the model.
- Wilks' Lambda is a positive-valued statistic that ranges from 0 to 1. Decreasing values of the statistic indicate effects that contribute more to the model.
- Hotelling's trace is the sum of the eigenvalues of the test matrix. It is a positive-valued statistic for which increasing values indicate effects that contribute more to the model. Hotelling's trace is always larger than Pillai's trace, but when the eigenvalues of the test matrix are small, these two statistics will be nearly equal. This indicates that the effect probably does not contribute much to the model.
- Roy's largest root is the largest eigenvalue of the test matrix. Thus, it is a positive-valued statistic for which increasing values indicate effects that contribute more to the model. Roy's largest root is always less than or equal to Hotelling's trace. When these two statistics are equal, the effect is predominantly associated with just one of the dependent variables, there is a strong correlation between the dependent variables, or the effect does not contribute much to the model.

drtamil@gmail.com

Which is better selection?

- There is evidence that Pillai's trace is more robust than the other statistics to violations of model assumptions.
- Each multivariate statistic is transformed into a test statistic with an approximate or exact F distribution. The hypothesis (numerator) and error (denominator) degrees of freedom for that F distribution are shown in the results.

drtamil@gmail.com

But some authors says;

- Specifically, if Box’s M is significant, then Pillai’s trace is preferred over the usual Wilks’ lambda. The larger the Pillai's trace, the more the given effect contributes to the model. Pillai's trace is always smaller than Hotelling's trace.
- MULTIVARIATE GLM, MANOVA, AND MANCOVA 2015 Edition by G. David Garson and Statistical Associates Publishing

drtamil@gmail.com

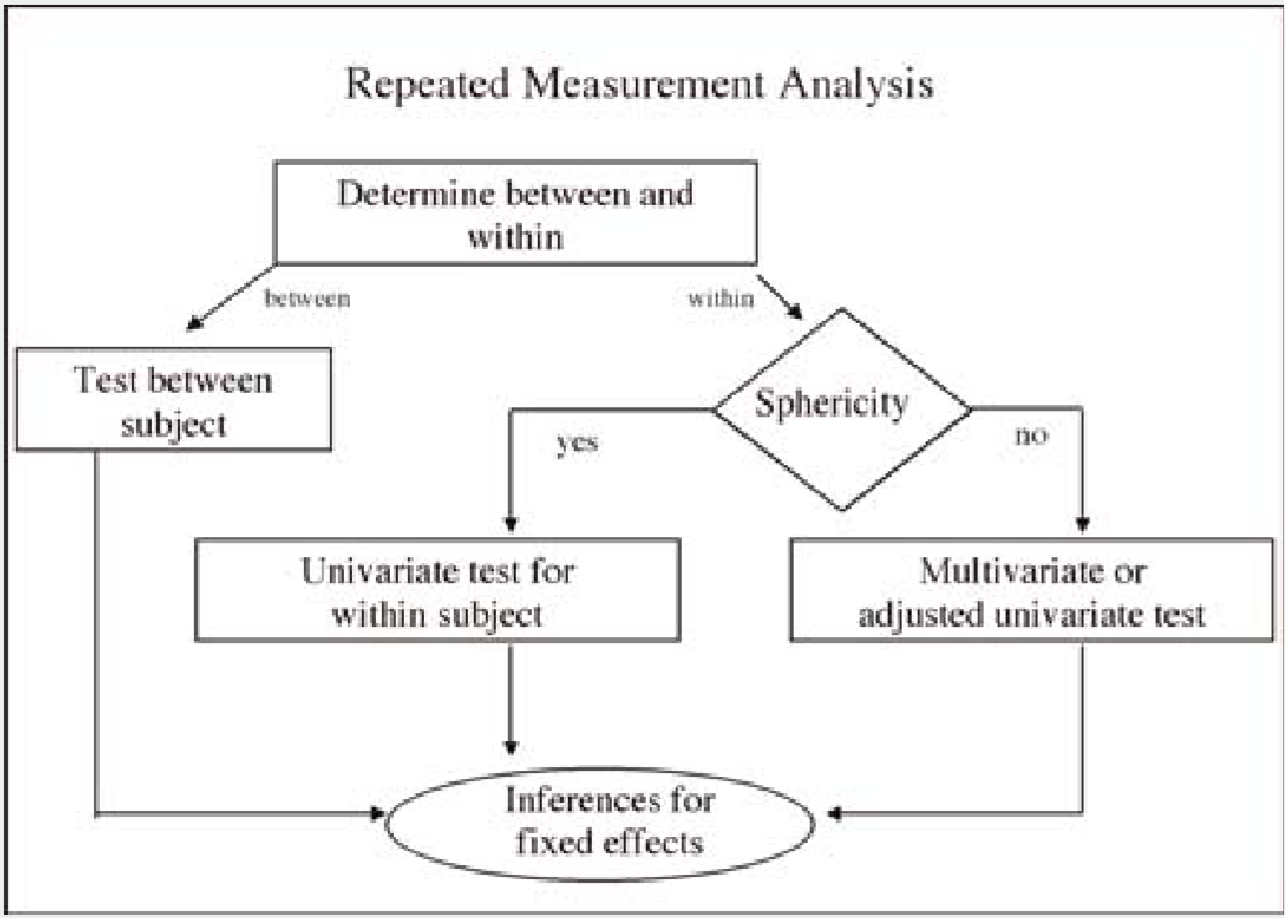
Univariate or Multivariate?

- Now both assumptions for Univariate and Multivariate procedures were valid. Which procedure should we use?
- Figure II gives the flowchart for the decision.
- Check the Sphericity assumption first- if satisfied, use the results from the Univariate procedure.
- Otherwise, proceed with the adjusted Univariate or Multivariate tests.

drtamil@gmail.com

Univariate or Multivariate?

Fig. 2 Flow chart for Repeated Measurement Analysis.



drtamil@gmail.com

What if the Mauchly Test of Sphericity is less than 0.05?

1. Proceed to Multivariate. Select appropriate test values after check Box Test.
2. If Box Test $p < 0.05$, then Multivariate homogeneity assumption is breached, proceed to adjusted Univariate.
3. Based on the following notes, decide on the appropriate value to be selected.

Dose

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	505037.863	1	505037.863	1662.542	.000	.960	1662.542	1.000
drug	91877.300	1	91877.300	302.452	.000	.814	302.452	1.000
Error	20960.447	69	303.775					

a. Computed using alpha = .05

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
dose1	.000	.	9	.	.353	.364	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + drug
Within Subjects Design: dose1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

the Mauchly's test does not provide any result if the sample size is less than the repeated measurement count.

So check box test, but no results. So must use adjusted univariate.

Warnings

Box's Test of Equality of Covariance Matrices is not computed because there are fewer than two nonsingular cell covariance matrices.
Post hoc tests are not performed for Drug because there are fewer than three groups.

drtamil@gmail.com

No **Mauchly's test** result if the sample size is less than the repeated measurement count.

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
dose1	.000	.	9	.	.353	.364	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + drug
Within Subjects Design: dose1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

- In SPSS, the Mauchly's test does not provide any result if the sample size is less than the repeated measurement count.

  Therefore, assume that the sphericity has been violated and go for a correction.

  Why? Epsilon estimates are measures of the deviation and values close to 1 indicate little or no problem, while values below 1 indicate progressively more severe departures. A sensible option is to rely on one of the corrections whenever epsilon estimates are below 0.9 or 0.95.

drtamil@gmail.com

Which epsilon?

- The degree to which sphericity is present, or not, is represented by a statistic called epsilon (ε).
- An epsilon of 1 (i.e., ε = 1) indicates that the condition of sphericity is exactly met. The further epsilon decreases below 1 (i.e., ε < 1), the greater the violation of sphericity. Therefore, you can think of epsilon as a statistic that describes the degree to which sphericity has been violated.
- The lowest value that epsilon (ε) can take is called the lower-bound estimate.
- Both the Greenhouse-Geisser and the Huynd-Feldt procedures attempt to estimate epsilon (ε), albeit in different ways (it is an estimate because we are dealing with samples, not populations). For this reason, the estimates of sphericity (ε) tend to always be different depending on which procedure is used.
- By estimating epsilon (ε), all these procedures then use their sphericity estimate (ε) to correct the degrees of freedom for the F-distribution.

drtamil@gmail.com

Greenhouse-Geisser Correction

The Greenhouse-Geisser procedure estimates epsilon (referred to as $\hat{\varepsilon}$) in order to correct the degrees of freedom of the *F*-distribution as has been mentioned previously, and shown below:

$$df_{time/condition} = \hat{\varepsilon}(k-1)$$
$$df_{error} = \hat{\varepsilon}(k-1)(n-1)$$

Using our prior example, and if sphericity had been violated, we would have:

$$df_{time/condition} = 0.638(3-1) = 1.276$$
$$df_{error} = 0.638(3-1)(6-1) = 6.380 \qquad where\ \hat{\varepsilon} = 0.638$$

So our *F*-test result is corrected from *F* (2,10) = 12.534, *p* = .002 to *F* (1.277,6.384) = 12.534, *p* = .009 (degrees of freedom are slightly different due to rounding). The correction has elicited a more accurate significance value. It has increased the *p*-value to compensate for the fact that the test is too liberal when sphericity is violated.

Huynd-Feldt Correction

As with the Greenhouse-Geisser correction, the Huynd-Feldt correction estimates epsilon (represented as $\tilde{\varepsilon}$) in order to correct the degrees of freedom of the *F*-distribution as shown below:

$$df_{time/condition} = \tilde{\varepsilon}(k-1)$$
$$df_{error} = \tilde{\varepsilon}(k-1)(n-1)$$

Using our prior example, and if sphericity had been violated, we would have:

$$df_{time/condition} = 0.760(3-1) = 1.520$$
$$df_{error} = 0.760(3-1)(6-1) = 7.600 \qquad where\ \tilde{\varepsilon} = 0.760$$

So our *F* test result is corrected from *F* (2,10) = 12.534, *p* = .002 to *F* (1.520,7.602) = 12.534, *p* = .005 (degrees of freedom are slightly different due to rounding). As with the Greenhouse-Geisser correction, this correction has elicited a more accurate significance value; it has increased the *p*-value to compensate for the fact that the test is too liberal when sphericity is violated.

drtamil@gmail.com

lower bound

- epsilon for lower bound = 0.250?
- Since have 5 comparisons, week=0, 1, 2, 4, 6.
- Epsilon lower bound = 1/(5-1) = 0.25.

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
dose1	.000	.	9	.	.353	.364	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + drug
Within Subjects Design: dose1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Greenhouse-Geisser vs. Huynd-Feldt Correction

The Greenhouse-Geisser correction tends to underestimate epsilon (ε) when epsilon (ε) is close to 1 (i.e., it is a conservative correction), whilst the Huynd-Feldt correction tends to overestimate epsilon (ε) (i.e., it is a more liberal correction). Generally, the recommendation is to use the Greenhouse-Geisser correction, especially if estimated epsilon (ε) is less than 0.75. However, some statisticians recommend using the Huynd-Feldt correction if estimated epsilon (ε) is greater than 0.75. In practice, both corrections produce very similar corrections, so if estimated epsilon (ε) is greater than 0.75, you can equally justify using either.

drtamil@gmail.com

So use Greenhouse Geisser correction since epsilon less than 0.75

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
dose1	.000	.	9	.	.353	.364	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + drug
Within Subjects Design: dose1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

drtamil@gmail.com

So use Greenhouse Geisser correction since epsilon less than 0.75

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
dose1	Sphericity Assumed	2794.830	4	698.707	8.840	.000	.114	35.359	.999
	Greenhouse-Geisser	2794.830	1.412	1978.809	8.840	.001	.114	12.485	.916
	Huynh-Feldt	2794.830	1.454	1922.038	8.840	.001	.114	12.854	.921
	Lower-bound	2794.830	1.000	2794.830	8.840	.004	.114	8.840	.834
dose1 * drug	Sphericity Assumed	349.759	4	87.440	1.106	.354	.016	4.425	.347
	Greenhouse-Geisser	349.759	1.412	247.638	1.106	.317	.016	1.562	.206
	Huynh-Feldt	349.759	1.454	240.534	1.106	.319	.016	1.609	.209
	Lower-bound	349.759	1.000	349.759	1.106	.297	.016	1.106	.179
Error(dose1)	Sphericity Assumed	21815.311	276	79.041					
	Greenhouse-Geisser	21815.311	97.454	223.852					
	Huynh-Feldt	21815.311	100.333	217.430					
	Lower-bound	21815.311	69.000	316.164					

a. Computed using alpha = .05

- Yes, there is a change over time (p=0.001). Dose is increasing over time.
- But no difference of change of dosage over time between the two drugs (p=0.317)

drtamil@gmail.com

Terima Kasih

MEAN(Trial1,Trial2,Trial3,Trial4)

drtamil@gmail.com

ANOVA

errors

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.521	1	2.521	.590	.460
Within Groups	42.729	10	4.273		
Total	45.250	11			

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	4800.000	1	4800.000	280.839	.000
Anxiety	10.083	1	10.083	.590	.460
Error	170.917	10	17.092		