

GLM Repeated-measures designs: One within-subjects factor

Reading: SPSS Advanced Models 9.0: 2. Repeated Measures

Homework: [Sums of Squares for Within-Subject Effects](#)

Download: [glm_within1.sav](#) ([Download Tips](#))

1. [Overview](#)
2. [The Data](#)
3. [Select the GLM Repeated Measures Procedure](#)
4. [How GLM Analyzes a Within-Subjects Factor](#)
5. [The Basic Output](#)
 - [Univariate Tests](#)
 - [Multivariate Tests](#)
6. [Interpreting Significant Effects: Displaying the Means](#)
7. [Interpreting Significant Effects: Post Hoc Tests](#)
8. [Alternative Contrasts for Within-Subject Factors](#)
9. [References](#)

Content

1. Overview

In a "within-subjects" design each participant provides more than one response. The "pre-post" design, where the participant responds both before the treatment and after the treatment, is a typical within-subjects design. The pre-post aspect of the design is a within-subjects factor. This type of design is also known as a "repeated measures" design.

This set of notes describes the GLM Repeated Measures procedures to run a repeated measures analysis of variance with one within-subjects factor.

top 

2. The Data

The data are from a treatment outcome study that looked at Eye Movement Desensitization and Reprocessing (EMDR) as a treatment for psychological trauma (Wilson, Becker, & Tinker, 1995, 1997). The 80 adult participants (40 males and 40 females) had experienced a traumatic event (e.g., physical-mental abuse, death of a significant other, health crisis) and were experiencing posttraumatic stress disorder (PTSD) symptoms such as flashbacks, nightmares, avoidant behaviors, increased anger). The traumatic event had occurred an average of 13.5 years prior to the beginning of the study. Forty-six percent of the participants met the DSM-IV criteria for PTSD, the remainder were classified as partial PTSD.

The participants were randomly assigned to an immediate treatment or a delayed treatment condition. Several measures of PTSD symptoms and general psychiatric symptoms were taken. In this set of notes we will look at the Intrusions scale of the Impact of Events Scale (IES; Horowitz, Wilner, & Alvarez, 1979) for the participants assigned to the immediate treatment condition. The IES Intrusions scale measures PTSD intrusion symptoms (e.g., nightmares, flashbacks). Assessments were made at four different times, see Table 1.

Table 1. Research Design for the Immediate Treatment Group

T1		T2	T3	T4
Pretreatment Assessment (<i>n</i> = 40)	Three 90 minute EMDR Treatment Sessions	Posttreatment Assessment (~30 days after T1) (<i>n</i> = 40)	Follow up Assessment at 3 Months Posttreatment (<i>n</i> = 40)	Follow up Assessment at 15 Months Posttreatment (<i>n</i> = 30)

The variables in the data file `glm_within1.sav` are shown in Table 2. The variable names are designed to include both the name of the scale (`iesi`) and the measurement time (e.g., `_t1`). Note that all four of the participants responses are contained on one record.

Table 2. The variables in `glm_within1.sav`

Variable Name	Variable Label / Value Label
<code>iesi_t1</code>	IES Intrusion, Pretreatment
<code>iesi_t2</code>	IES Intrusion, Posttreatment
<code>iesi_t3</code>	IES Intrusion, 3-month follow-up
<code>iesi_t4</code>	IES Intrusion, 15-month follow-up

top 

3. Select the GLM Repeated Measures Procedure

The GLM Repeated Measures dialog box is opened by clicking

Analyze

General Linear Model

Repeated Measures...

Repeated Measures Define Factor(s)

The within-subject factors are defined in the opening window. A within-subject factor is defined by a name, entered in the **Within-Subject Factor Name:** window and by the number of levels, entered into the **Number of Levels:** window. The naming conventions are the same as for any SPSS variable name. It should start with a letter and it can be no longer than eight characters long. Lets call the within-subjects factor in this example *time*. The number of levels of the within subject factor corresponds to the number of measures for the factor. We have four measures, `iesi_t1`, `iesi_t2`, `iesi_t3`, and `iesi_t4`, so enter the number 4 in the Number of Levels: box. Press the **Add** button to add *time(4)* to the list of within subject factors. In summary, we have created a within subjects factor out of the four IES intrusion scores (`iesi` at `t1`, `t2`, `t3`, and `t4`). The name of the within

subjects factor is *time* with four levels.

Repeated Measures

The next step is to Define each of the four levels of the time within-subjects variable. Move each of the variables to the **Within-Subjects Variables (time)**: window. Make sure that the conceptual order of the original variables is preserved when you move them to the within-subjects variables window, that is, *iesi_t1* should be moved to the __?__(1) space, *iesi_t2* should be moved to the __?__(2) space, and so forth. The variables in the data editor are in the correct order so you could highlight all four variables and move them to **Within-Subjects Variables** window all at once.

This design has no **between-subjects factors** or **covariates** so those windows remain empty.

top 

4. How GLM Analyzes a Within-Subjects Factor

In order to compute a within-subjects effect, GLM **transforms** the within-subject variables into a new set of variables, one variable for each degree of freedom of the within-subject variable plus one additional variable for the average of the within-subject factor. The analysis of variance is performed on the transformed variables rather than on the original within-subject variables.

By default, GLM uses a set of orthogonal *polynomial* transformations. The three polynomial transformations in this example represent the *linear*, *quadratic*, and *cubic effects* of time. Each of the new variables (linear, quadratic, and cubic) is a separate estimate of the main effect.

You can choose other sets of contrasts for within-subject factors by going to the **Contrasts...** dialog box and selecting an alternative contrast. For example the "repeated" contrast will create the following set of 1 df contrasts for the time main effect: level 1 vs. level 2, level 2 vs. level 3, level 3 vs. level 4. In this example the contrasts would be: *iesi_t1* vs. *iesi_t2*, *iesi_t2* vs. *iesi_t3*, and *iesi_t3* vs. *iesi_t4*.

The analysis is done on the transformed variables (*linear*, *quadratic*, and *cubic*) rather than on the original variables (*iesi_t1* to *iesi_t4*). The new transformed variables are computed by assigning weights (also called coefficients) to each of the four scores, and then adding those weighted scores together.

The matrix of coefficients that are used to create the transformed variables are printed when you check **Transformation Matrix** in the **Options...** dialog box. The coefficients for the transformed scores are shown in Tables 3 and 4.

Table 3. Average
Measure: MEASURE_1
Transformed Variable: AVERAGE

IES Intrusion, pretreatment	.500
IES Intrusion, posttreatment	.500
IES Intrusion, 3 months posttreatment	.500
IES Intrusion, 15 months post treatment	.500

Table 4. TIME(a)
Measure: MEASURE_1

Dependent Variable	TIME		
	Linear	Quadratic	Cubic
IES Intrusion, pretreatment	-.671	.500	-.224
IES Intrusion, posttreatment	-.224	-.500	.671
IES Intrusion, 3 months posttreatment	.224	-.500	-.671
IES Intrusion, 15 months post treatment	.671	.500	.224

a The contrasts for the within subjects factors are:
TIME: Polynomial contrast

The new transformed variable *AVERAGE* is computed by multiplying each of the original variables by its coefficient. The coefficients are .500 for each variable so the variable *average* is computed using the following formula

$$\text{average} = (.5)\text{iesi_t1} + (.5)\text{iesi_t2} + (.5)\text{iesi_t3} + (.5)\text{iesi_t4}$$

The transformed variable *average* would be used to test main effects and interactions for any between-subjects factors. The weights are chosen so that:

- 1) the contribution of each variable to the average score is the same (the weights are the same).
- 2) the sum of the squared weights is equal to 1. In this example,

$$.5^2 + .5^2 + .5^2 + .5^2 = 1.00$$

By default, GLM will create three orthogonal polynomial transformations from the four original measures. The coefficients are chosen so that:

- 1) The sum of the coefficients = 0.
- 2) The sum of the squared coefficients = 1.00. This assures that each transformation receives the same weight in the analysis. This is called *normalizing* the transformation.
- 3) The sum of the cross products of any two sets of coefficients = 0. This assures that each transformation is orthogonal to other transformations.

The three transformed variables, *Linear*, *Quadratic*, and *Cubic* are found by multiplying the original variables by their respective coefficients, see Table 4.

$$\text{Linear} = (-.671*\text{iesi_t1}) + (-.224*\text{iesi_t2}) + (.224*\text{iesi_t3}) + (.671*\text{iesi_t4})$$

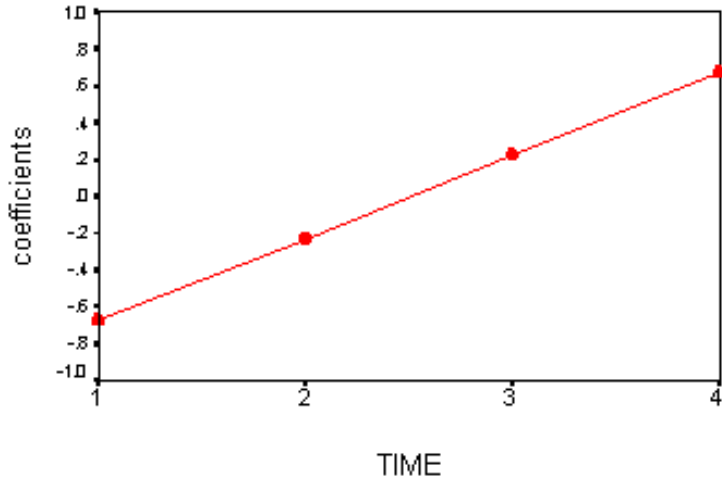
$$\text{Quadratic} = (.500*\text{iesi_t1}) + (-.500*\text{iesi_t2}) + (-.500*\text{iesi_t3}) + (.500*\text{iesi_t4})$$

$$\text{Cubic} = (-.224*\text{iesi_t1}) + (.671*\text{iesi_t2}) + (-.671*\text{iesi_t3}) + (.224*\text{iesi_t4})$$

The plots of the linear, quadratic, and cubic coefficients are shown below. Notice that the plots of the coefficients look like the underlying pattern (linear, quadratic, cubic) that they were designed to detect.

Orthogonal Polynomial Coefficients:

Linear Trend

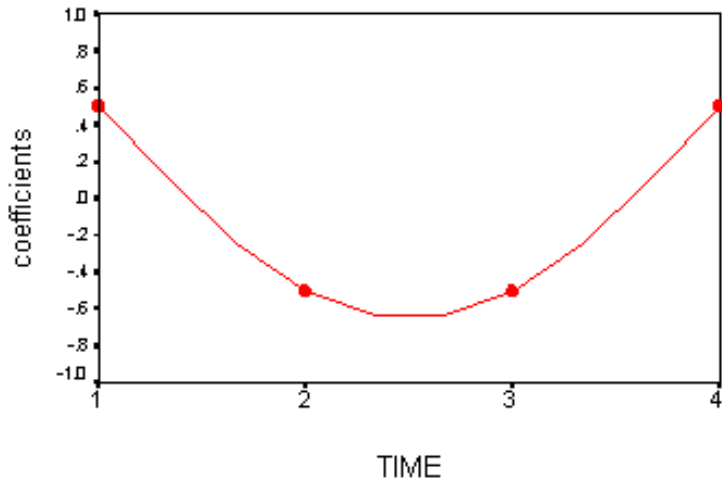


The null hypothesis of no linear trend is that the trend = 0. In this example,

$$\begin{aligned} \text{Linear} = & (-.671*iesi_t1) + \\ & (-.224*iesi_t2) + \\ & (.224*iesi_t3) + \\ & (.671*iesi_t4) \\ & = 0 \end{aligned}$$

Orthogonal Polynomial Coefficients:

Quadratic Trend

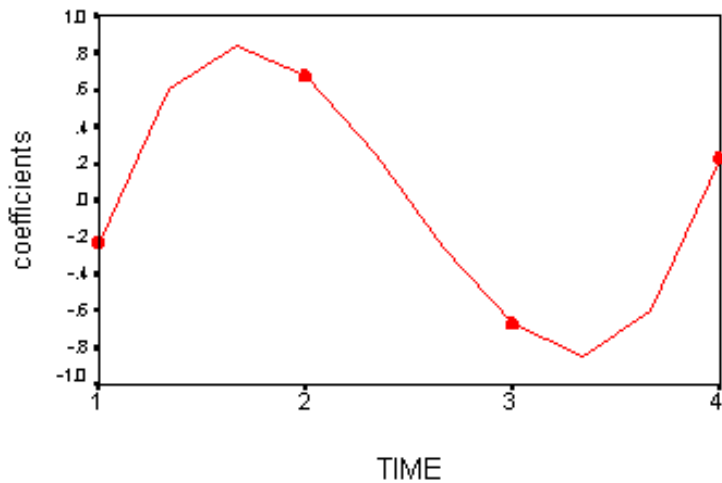


In this example, the null hypothesis of no quadratic trend is --

$$\begin{aligned} \text{Quadratic} = & (.500*iesi_t1) + \\ & (-.500*iesi_t2) + \\ & (-.500*iesi_t3) + \\ & (.500*iesi_t4) \\ & = 0. \end{aligned}$$

Orthogonal Polynomial Coefficients:

Cubic Trend



In this example, the null hypothesis of no cubic trend is --

$$\begin{aligned} \text{Cubic} = & (-.224*iesi_t1) + \\ & (.671*iesi_t2) + \\ & (-.671*iesi_t3) + \\ & (.224*iesi_t4) \\ & = 0. \end{aligned}$$

Let use the following syntax commands to compute the linear, quadratic, and cubic trends --

```
COMPUTE AVERAGE = (.500*iesi_t1) + (.500*iesi_t2) + (.500*iesi_t3) + (.500*iesi_t4).
COMPUTE LINEAR = (-.671*iesi_t1) + (-.224*iesi_t2) + (.224*iesi_t3) + (.671*iesi_t4).
COMPUTE QUAD = (.500*iesi_t1) + (-.500*iesi_t2) + (-.500*iesi_t3) + (.500*iesi_t4).
COMPUTE CUBIC = (-.224*iesi_t1) + (.671*iesi_t2) + (-.671*iesi_t3) + (.224*iesi_t4).
EXECUTE.
```

Find the linear, quadratic, and cubic trends for the following sets of scores (You can do this by adding these values to the end of the glm_with1.sav data file used in this example and then running the above set of syntax commands).

iesi_t1	iesi_t2	iesi_t3	iesi_t4	linear	quad	cubic
1	1	1	1			
15	15	15	15			
2	4	6	8			
8	6	4	2			
10	6	6	10			
6	10	10	6			
6	1	8	4			

SPSS actually computes the values of the transformed variables (linear, quadratic, and cubic) for each person. The transformed variables are **temporary** variables that only exist while the GLM procedure is computing the analysis. When the computations are complete the variables no longer exist. You never see them in the Data Editor.

See notes below on [Alternative contrasts](#) for additional contrasts.

top 

5. The Basic Output for GLM Repeated Measures

The basic output for GLM repeated measures includes the following tables: Within Subjects Factors (see Table 5); Mauchley's Test of Sphericity (see Table 6); Tests of Within-Subjects Effects (see Table 7); Tests of Within-Subjects Contrasts (see Table 8), and Tests of Between-Subjects Effects (none in this analysis), and Multivariate Tests (see Table 9).

Within-Subject Factor Definition

Table 5. Within-Subjects Factors

Measure: MEASURE_1

TIME	Dependent Variable
1	IESI_T1
2	IESI_T2
3	IESI_T3
4	IESI_T4

Table 5 shows the four levels of the within-subject factor called *time*. The name of the dependent variable is shown for each of the four levels.

The default name of the measure is *measure_1*. We could have given our own name for the measure, perhaps *IES_I*. The measure can be named in the original repeated measures dialog box, press the **Measures>>** button to open the name measure part of the dialog box.

Univariate Tests

Univariate Assumptions

The within-subject effect is analyzed in GLM by first transforming the original variables into single degree of freedom tests of the null hypothesis. In our example the *time* within-subject effect has four levels so there are three degrees of freedom for the time main effect and GLM will create three new variables called linear, quadratic, and cubic. In order to create the overall test of the time main effect with 3 df, GLM will add together the three single degree of freedom main effect estimates. That is, it will sum the sums of squares for each of the three new variables. It is reasonable to add those variables together only if they meet two conditions:

- (a) their variances are equal, and
- (b) they are uncorrelated with each other.

Equal variances. The within-subjects error term is found by summing the error sums of squares for each of the transformed variables. The error variances of each of the transformed scores should be homogeneous.

Correlation between transformed scores. The sums of squares for the within-subjects main effect are found by summing the sums of squares for the new transformed variables. If there is a significant correlation between the transformed scores then adding together the sums of squares will yield an overestimate of the amount variance due to the method effect because the correlated variance will be counted twice. It may be useful to think of a Venn diagram. If the *linear* and *quadratic* variables are uncorrelated then the circles representing the variances of *linear* and *quadratic* are nonoverlapping. Adding together their variances provides an accurate estimate of the total amount of variance accounted for by the two transformed scores. However, if the *linear* and *quadratic* variables are correlated, then the circles representing the variances of *linear* and *quadratic* will overlap. The overlapping part is the amount of shared variance. When you add together the variances of *linear* and *quadratic* the shared variance is counted twice, giving an overestimate of the amount of variance accounted for by those variables.

The Mauchly test of sphericity tests both of those assumptions at the same time. If the Mauchly test is not significant, then it is appropriate to add the three single degree of freedom estimates together to get the overall estimate with three degrees of freedom.

If the sphericity assumption is not met the averaged F-tests overestimate the strength of the relationships. If you have a significant averaged F-test in the analysis of variance you have two options. One option is ignore the averaged F-tests, and report the multivariate test of significance. The other option is to apply a correction to the averaged F-tests. The correction involves multiplying both the effect df and the error df by one of the Epsilons provided for you. The Huynh-Feldt Epsilon is one of the more commonly used correction formulas.

The Mauchly test is automatically included in all output for repeated measures designs, see Table 6. In this example the

Mauchly's W is significant, $W(5) = 0.232$, $p < .0005$, so the sphericity assumption has not been met. You should use one of the epsilons to correct both degrees of freedom for the time main effect and then look up the F value using the new degrees of freedom. The correction is made by multiplying the df for the effect by the epsilon value and by multiplying the df for the error term by the epsilon value. The significance of the F value is then determined using the corrected df . As note "b" in Table 6 states, the corrected tests are displayed in the layers (by default) of the Tests of Within Subjects Effects table.

Table 6. Mauchly's Test of Sphericity(b)

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon(a)		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	.232	40.478	5	.000	.561	.592	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized 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 layers (by default) of the Tests of Within Subjects Effects table.

b Design: Intercept
Within Subjects Design: TIME

Univariate ANOVA Results

Univariate tests of the *time* within-subject effect are shown in Table 7. The **sphericity assumed** row displays the unadjusted statistics. It is appropriate if the sphericity assumption has been met. The next three rows report the statistics after one of the epsilons has been applied to the degrees of freedom. One of the epsilon corrected tests should be used when the sphericity assumption has not been met.

Table 7. Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
TIME	Sphericity Assumed	4157.500	3	1385.833	62.665	.000	.684
	Greenhouse-Geisser	4157.500	1.684	2469.524	62.665	.000	.684
	Huynh-Feldt	4157.500	1.776	2341.339	62.665	.000	.684
	Lower-bound	4157.500	1.000	4157.500	62.665	.000	.684
Error(TIME)	Sphericity Assumed	1924.000	87	22.115			
	Greenhouse-Geisser	1924.000	48.822	39.408			
	Huynh-Feldt	1924.000	51.495	37.363			
	Lower-bound	1924.000	29.000	66.345			

The corrected statistics were found by multiplying both effect and error df by the epsilon value.

The df after applying the Greenhouse-Geisser epsilon, .561, are

$$\begin{aligned}\text{corrected df effect} &= \text{original df for effect} * \text{Greenhouse-Geisser epsilon} \\ &= 3 * .561 \\ &= 1.68\end{aligned}$$

$$\begin{aligned}\text{corrected df error} &= \text{original df for error} * \text{Greenhouse-Geisser epsilon} \\ &= 87 * .561 \\ &= 48.81\end{aligned}$$

Similarly the df after applying the Huynh-Feldt epsilon, .592, are

$$\begin{aligned}\text{corrected df effect} &= \text{original df for effect} * \text{Huynh-Feldt epsilon} \\ &= 3 * .592 \\ &= 1.78\end{aligned}$$

$$\begin{aligned}\text{corrected df error} &= \text{original df for error} * \text{Huynh-Feldt epsilon} \\ &= 87 * .592 \\ &= 51.50\end{aligned}$$

The df for the Lower bound statistics were found in the same manner. Lower bound df are rounded to the nearest integer. The lower bound epsilon is very conservative, it is rarely used.

The values of epsilons are always less than or equal to 1. The smaller the epsilon the smaller the corrected df, and consequently the larger the F value needs to be to be significant. Because the epsilons are less than or equal to 1.00 they can be thought of as percents. Correcting a df by an epsilon of, say, .69 will result in a corrected df that is 69% of the original df.

Using the Huynh-Feldt correction the main effect time was significant, $F(1.78, 51.50) = 62.66, p < .0005$, partial Eta squared = .684. The next step would be to describe those differences.

Note that the partial Eta squared statistic does not change when the various epsilon values are applied to the analysis. Partial Eta squared is a function of the Sums of Squares for the effect and the error (see [Statistical Note: Effect Size](#))

Univariate Single df Tests (Tests of Contrasts)

The sums of squares and df for the univariate test of the *time* within-subjects effect shown in Table 7 were found by summing the sums of squares and df for the single degree of freedom tests of the *time* effect (see Table 8).

$$\begin{aligned}SS_{\text{time}} &= SS_{\text{linear}} + SS_{\text{quadratic}} + SS_{\text{cubic}} \\ 4157.500 &= 2974.827 + 952.033 + 230.640\end{aligned}$$

The analysis of the transformed variables shows that each effect was significant at $p = .001$ or less. The default, orthogonal polynomial transformations may or may not be useful to you in describing the pattern of the means. In this study they are not very helpful.

Table 8. Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
	Linear	2974.827	1	2974.827	106.892	.000	.787
	Quadratic	952.033	1	952.033	41.866	.000	.591
	Cubic	230.640	1	230.640	14.621	.001	.335
Error(TIME)	Linear	807.073	29	27.830			
	Quadratic	659.467	29	22.740			
	Cubic	457.460	29	15.774			

Multivariate Tests

Multivariate Assumptions

The multivariate assumption is that the measurements are sampled from a multivariate normal distribution. There are no additional assumptions if there are no between-subjects factors. Because the sphericity assumption does not apply to the multivariate tests, the multivariate analysis is often viewed as an alternative to the univariate repeated measures ANOVA when the sphericity assumption has not been met.

Multivariate ANOVA Results

GLM repeated measures always creates a multivariate test of the hypothesis, see Table 9. Four different multivariate tests are reported, Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root. A description of the differences between these measures is beyond this set of notes. The F ratio's associated with the multivariate tests are all the same for this analysis.

The *time* effect was significant, multivariate $F(3, 27) = 38.30, p < .0005$, partial Eta squared = .810.

Table 9. Multivariate Tests^b

Effect		Value	F	Hypothesis df	Error df	Sig.	Eta Squared
TIME	Pillai's Trace	.810	38.304 ^a	3.000	27.000	.000	.810
	Wilks' Lambda	.190	38.304 ^a	3.000	27.000	.000	.810
	Hotelling's Trace	4.256	38.304 ^a	3.000	27.000	.000	.810
	Roy's Largest Root	4.256	38.304 ^a	3.000	27.000	.000	.810
a Exact statistic							
b Design: Intercept Within Subjects Design: TIME							

Tests of Between Subject Effects

There are no between-subjects effects in this design. The table by that name in the Output Navigator lists a test of the *intercept*. It is a test of the hypothesis that the transformed variable *average* is not significantly different from zero. That is, is the average of

the original variables is not different from zero. Because all the IES scores are above zero the test will always be significant. This test is not relevant.

top 

6. Interpreting Significant Effects: Displaying the Means

There are three different ways to display the means: (a) display the descriptive statistics; (b) display the estimated marginal means, and (c) display profile plots.

Descriptive Statistics

Descriptive statistics are displayed by checking the **Display Descriptives** box in the **Options...** dialog box.

The descriptive statistics for each of the four variables (iesi_t1, iesi_t2, iesi_t3, and iesi_t4) are shown in Table 10.

Note that there are equal *ns* for each measure. **For a within-subjects factor missing values are deleted on a casewise basis.** That is, if any of the scores are missing for a variable then the entire case is deleted from the analysis. In this study 30 of the 40 participants were available for retesting at the 15 month follow-up.

Table 10. Descriptive Statistics

	Mean	Std. Deviation	N
IES Intrusion, pretreatment	18.57	8.29	30
IES Intrusion, posttreatment	6.00	7.31	30
IES Intrusion, 3 months posttreatment	5.27	6.88	30
IES Intrusion, 15 months post treatment	3.97	5.93	30

Estimated Marginal Means

To display the estimated marginal means go to the **Options...** dialog box and move the *time* factor to the **Display Means for...** window.

The estimated marginal means, standard errors, and 95% confidence intervals are shown in Table 11.

Table 11. Estimates

Measure: MEASURE_1

TIME	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	18.567	1.513	15.472	21.661
2	6.000	1.334	3.272	8.728
3	5.267	1.257	2.697	7.837
4	3.967	1.083	1.751	6.182

Profile Plots

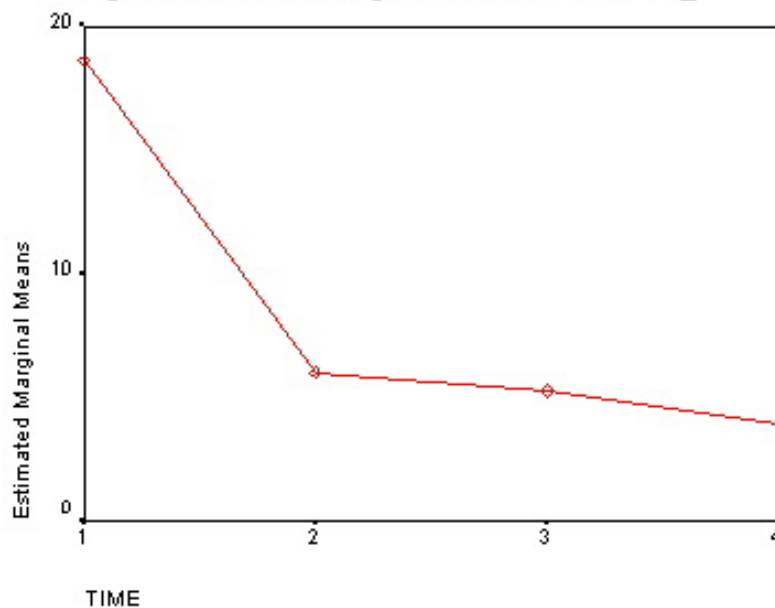
To display a profile plot:

- (a) go to the **Plots...** dialog box,
- (b) move the *time* factor to the **Horizontal Axis** window,
- (c) **Add** the plot to the Plots window, and
- (d) press **Continue**.

The profile plot is shown in Figure 1.

The plot suggests that there is a large drop in IES intrusion symptoms from the pretreatment to the posttreatment assessment periods and that the changes that occur at the 3-month and 15-month follow-up are small relative to the change associated with the treatment. There doesn't seem to be any increase in symptoms at the 15 month follow-up. If anything there may be a slight improvement in the symptoms when compared with the posttreatment mean.

Figure 1. Estimated Marginal Means of MEASURE_1

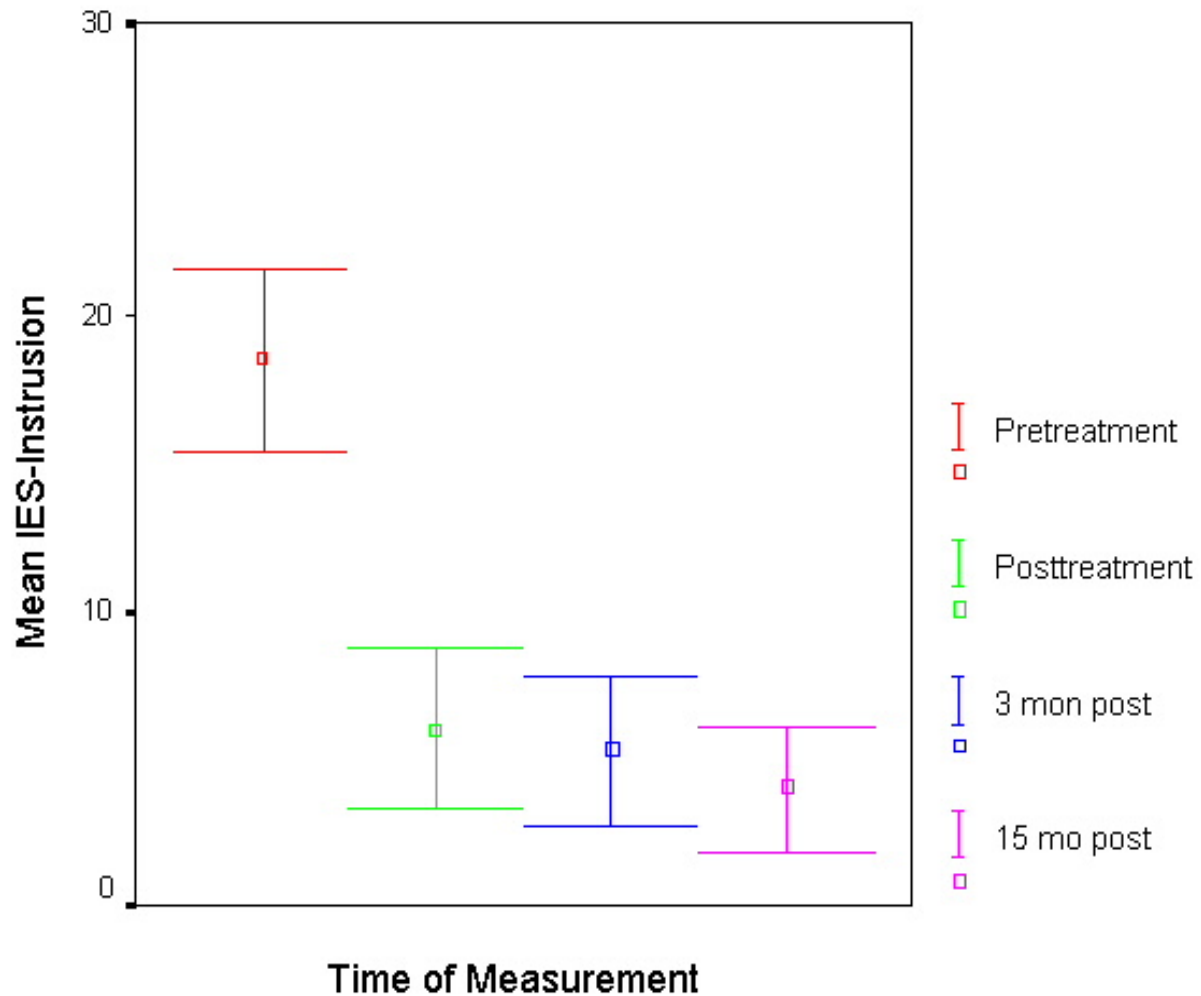


Plots with Error Bars

The error bars produced by SPSS graphics are based on the standard errors of the individual means. However, tests of differences between within-subject means are based on the standard errors of the differences between each pair of means. The standard error of the difference is always smaller than the standard error of the individual means if the two scores are correlated with each other. Because the two within-subject scores are based on the same participants it is nearly always true that scores are correlated. Hence the 95% confidence intervals for the error bar plots are too wide. They are biased towards a Type II error, incorrectly failing to reject the null hypothesis.

The error bar plot in Figure 2 shows that there was significant improvement from pretreatment to posttreatment. The improvement found at the posttest was maintained at both the 3- and 15-month follow-up times. But there were no differences between any of the post-treatment scores. Compare that interpretation with that provided by the Bonferroni-corrected paired comparison tests described later in this set of notes.

Figure 2. Estimated means with 95% confidence interval error bars



This plot was created from the Graphics - Error Bar... program (not from Interactive Graphics) using the Clustered and Data in Chart Are Summaries of separate variables options. The four scores (pretest, posttest, 6 month and 15 month follow-up) were moved to the Variables: window. A category axis is required so a dummy category axis was created by computing a new variable, valid, that had a single constant, 1, as its value. That new variable was used as the Category Axis: variable. The headings and legends were edited for content and font size.

top 

7. Interpreting Significant Effects: Post Hoc Pairwise Comparisons

A. Recommend Method: Compare Main Effects Using Estimated Marginal Means

The steps for making paired comparisons among the *time* within-subject effect means are as follows::

- go to the **Estimated Marginal Means** section of the **Options...** dialog box;
- move the *time* factor from the **Factors(s) and Factor Interaction** window to the **Display Means for...** window;
- check the **Compare main effects** box;
- select an adjustment for the confidence interval (e.g., Bonferroni), and

(d) press continue.

The paired comparisons are shown in Table 12.

Table 12. Pairwise Comparisons

Measure: MEASURE_1

(I) TIME	(J) TIME	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	12.567*	1.588	.000	8.071	17.063
	3	13.300*	1.579	.000	8.830	17.770
	4	14.600*	1.448	.000	10.501	18.699
2	1	-12.567*	1.588	.000	-17.063	-8.071
	3	.733	.901	1.000	-1.817	3.284
	4	2.033*	.539	.004	.506	3.560
3	1	-13.300*	1.579	.000	-17.770	-8.830
	2	-.733	.901	1.000	-3.284	1.817
	4	1.300	.797	.681	-.956	3.556
4	1	-14.600*	1.448	.000	-18.699	-10.501
	2	-2.033*	.539	.004	-3.560	-.506
	3	-1.300	.797	.681	-3.556	.956

Based on Estimated Marginal Means
 *. The mean difference is significant at the .05 level.
 a. Adjustment for multiple comparisons: Bonferroni

Writing up the Results of the ANOVA and the Pairwise Comparisons

Begin the write-up by describing the analysis of variance.

The intrusion score of the Impact of Events Scale (IES-I) was analyzed in an analysis of variance with time of measurement (pretest vs. immediate posttest vs. 3-month follow-up vs. 15-month follow-up) as a within-subjects factor.

Note that the sphericity assumption was not met and specify the correction that was used in the analysis. Then describe the results of that analysis.

The intrusion score of the Impact of Events Scale (IES-I) was analyzed in an analysis of variance with time of measurement (pretest vs. immediate posttest vs. 3-month follow-up vs. 15-month followup) as a within-subjects factor. **The sphericity assumption was not met so the Huynh-Feldt correction was applied. The main effect of time of measurement was significant, $F(1.78, 51.50) = 62.67, p < .0005, \eta^2 = .68$.**

Then describe the post-hoc tests and the results of those tests.

The intrusion score of the Impact of Events Scale (IES-I) was analyzed in an analysis of variance with time of measurement (pretest vs. immediate posttest vs. 3-month follow-up vs. 15-month follow-up) as a within-subjects factor. The sphericity assumption was not met so the Huynh-Feldt correction was applied. The main effect of time of measurement was significant, $F(1.78, 51.50) = 62.67$, $p < .0005$, $\eta^2 = .68$. **Post-hoc comparisons were performed using the Bonferroni adjustment for multiple comparisons. The EMDR treatment was effective in reducing PTSD intrusion symptoms. The IES-I score was reduced from a mean of 18.57 (SD = 8.29) at pretreatment to a mean of 6.00 (SD = 7.31, $p < .0005$) immediately following treatment. The improvement was maintained at the 3-month (M = 5.27, SD = 6.88, $p < .0005$) and 15-month (M = 3.97, SD = 5.93, $p < .0005$) follow-ups. There was no difference between the immediate posttreatment mean and the 3-month follow up mean ($p = 1.00$). The mean IES-I score at the 15-month follow-up was not different from the 3-month follow-up mean ($p = .681$), but it was significantly lower than the immediate posttreatment mean ($p = .004$).**

In this example, the Sidak comparison gives identical paired-comparison outcomes.

Note that the interpretation based on the Bonferroni corrected pairwise comparisons is different from the interpretation you might have made from looking at the profile plot with the 95% confidence intervals in Figure 2. As discussed above, the confidence intervals in that profile plot are too conservative.

Statistical Note: Computation of the 95% Confidence Intervals

These within-subject pairwise comparisons can be thought of as a series of paired t tests. The standard error is based on the standard deviation of each pair of difference scores. The 95% confidence interval is found as

$$95\% \text{ C.I.} = q_{\alpha,p,v} * \text{Standard Error of the Difference}$$

where $q_{\alpha,p,v}$ is obtained from the Percentage Points for the Studentized Range Statistic table at a given significance level, α , with p means, and v degrees of freedom for the Standard error of the difference term. The degrees of freedom, v , are the number of cases in the paired comparison - 1 --

$$\begin{aligned} \text{df} &= n \text{ of paired cases} - 1 \\ &= 30 - 1 \\ &= 29 \end{aligned}$$

The significance level may be corrected for the LSD test (no correction), a Bonferroni test, or a Sidak test. The Bonferroni correction is--

$$\text{Bonferroni alpha} = \alpha / C$$

where C is the number of possible paired comparisons. In this example with 4 means, there are 6 possible comparisons, so the Bonferroni alpha is

$$\text{Bonferroni alpha} = .05 / 6 = .008333 .$$

You would look up the q value with an alpha level of .0083, 2 means, and 29 degrees of freedom. Of course we don't have tables with exact probability levels so it is difficult to "look up" the value of q . SPSS computes the exact value of q from these

values.

The Sidak correction is

$$\text{Sidak alpha} = 1 - (1 - \alpha)^{1/C}$$

where C is the number of possible paired comparisons. In this example with 4 means the Sidak corrected alpha is

$$\begin{aligned} \text{Sidak alpha} &= 1 - (1 - .05)^{1/6} \\ &= 1 - ((0.95) * 0.16667) \\ &= 1 - 0.991488 \\ &= 0.008512 \end{aligned}$$

Note that there is no "homogeneous subset" output. Each comparison is computed using the standard deviation of the difference scores for that comparison.

B. Post Hoc... Comparisons

The comparisons in the **Post Hoc...** dialog box are not available for within-subject effects.

C. Within-Subject Main Effects: Running Tukey's HSD Test by Hand

You could run a Tukey post-hoc test by hand. But given that you can run the Bonferroni test using the Compare Means option, why would you want to? If you insist (my mentor made me do it), here are the guidelines to follow...

Recall that the formula for Tukey's HSD is as follows:

$$\psi(HSD) = q_{\alpha, p, v} \sqrt{\frac{MS_{\text{error}}}{n}}$$

You can compute the HSD using the appropriate MSerror for the univariate within-subjects main effect. But what is the appropriate MSerror? The ANOVA table, reproduced below, provides three error terms for the Time main effect labeled: Sphericity Assumed, Greenhouse-Geisser, Huynh-Feldt, and Lower-Bound.

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	4157.500	3	1385.833	62.665	.000
	Greenhouse-Geisser	4157.500	1.684	2469.524	62.665	.000
	Huynh-Feldt	4157.500	1.776	2341.339	62.665	.000
	Lower-bound	4157.500	1.000	4157.500	62.665	.000

Error(TIME)	Sphericity Assumed	1924.000	87	22.115		
	Greenhouse-Geisser	1924.000	48.822	39.408		
	Huynh-Feldt	1924.000	51.495	37.363		
	Lower-bound	1924.000	29.000	66.345		

If there is no sphericity problem, that is, if Mauchly's W is not significant, then you should use the Sphericity Assumed MSerror. If there is a sphericity problem then you should use the MSerror for whichever epsilon correction use decided to use to test the main effect.

For this example the Mauchly's test of sphericity was significant, Mauchly's $W(5) = 0.232, p < .005$. Suppose that you decided to use the Huynh-Feldt correction, then the appropriate MSerror for the HSD test would be 37.363 with 51.495 degrees of freedom for the error term.

The value of q at $\alpha = .05, p$ means = 4, and $df = 51.495$ (the epsilon corrected degrees of freedom) is approximately 3.76. The critical value for the HSD test would then be

$$\begin{aligned} \text{HSD} &= 3.76 * \text{SQRT}(37.363/30) \\ &= 3.76 * 1.1160 \\ &= 4.196 \end{aligned}$$

If the critical difference between two means is equal to or greater than 4.196, then those means are different at the .05 level of significance. The mean differences are shown at the right.

In our example the pretest scores are greater than each of the posttest and follow-up scores, and none of the posttest and follow-up scores are different from each other.

		Mean Difference (I-J)
(I) TIME(Mean)	(J) TIME(Mean)	
1(18.567)	2(6.000)	12.567*
	3(5.267)	13.300*
	4(3.967)	14.600*
2(6.000)	1(18.57)	-12.567*
	3(5.267)	.733
	4(3.967)	2.033
3(5.267)	1(18.567)	-13.300*
	2(6.000)	-.733
	4(3.967)	1.300
4(3.967)	1(18.567)	-14.600*
	2(6.000)	-2.033
	3(5.267)	-1.300

* $p < .05$ using Tukey's HSD

Writing up the Results of the ANOVA and the Tukey Post Hoc Tests

Begin the write-up by describing the analysis of variance.

The intrusion score of the Impact of Events Scale (IES-I) was analyzed in an analysis of variance with time of measurement (pretest vs. immediate posttest vs. 3-month follow-up vs. 15-month follow-up) as a within-subjects factor. The sphericity assumption was not met so the Huynh-Feldt correction was applied. The main effect of time of measurement was significant, $F(1.78, 51.50) = 62.67, p < .0005, \eta^2 = .68$.

Then describe the type of post hoc test that was used and the results of the paired comparisons. A statement that summarizes the results of the post-hoc comparisons is often helpful.

The intrusion score of the Impact of Events Scale (IES-I) was analyzed in an analysis of variance with time of measurement (pretest vs. immediate posttest vs. 3-month follow-up vs. 15-month followup) as a within-subjects factor. The sphericity assumption was not met so the Huynh-Feldt correction was applied. The main effect of time of measurement was significant, $F(1.78, 51.50) = 62.67, p < .0005, \eta^2 = .68$. **Post hoc paired comparisons were made using Tukey's HSD test with p set at .05. The Huynh-Feldt corrected mean square error and degrees of freedom were used in calculating the HSD critical value. There was a decrease in the IES-I scores from pretest ($M = 18.57, SD = 8.29$) to immediate posttest ($M = 6.00, SD = 7.31$). The IES-I means at the 3-month follow-up ($M = 5.27, SD = 6.88$) and the 15-month follow-up ($M = 3.97, SD = 5.93$) were lower than the pretest mean, but they not different from each other or the immediate posttest mean. These results indicate that the improvement shown at the immediate posttest was maintained across the 3- and 15-month follow-up periods.**

A caution:

When there is a sphericity problem this approach to paired comparison testing provides a somewhat conservative standard error for all comparisons. A better approach would be to run paired comparisons using the **compare main effects** option in the **estimated marginal means** section of the **Options...** dialog box. Those tests are based on the standard errors of the differences of the individual paired comparisons so no corrections for sphericity are necessary.

top 

8. Alternative Contrasts for Within-Subject Factors

By default, single degree of freedom contrasts for within subject factors in GLM are orthogonal polynomial contrasts. Those contrasts may (or may not) be useful if you have a time based repeated measure. If you do not have a time based repeated measure they will probably not be helpful to you in interpreting a significant main effect with more than 1 degree of freedom. Several other contrasts are available including: deviation, simple, difference, Helmert, and repeated. [Choosing the option of "none," prints the default orthogonal polynomial single degree of freedom contrasts.

The selection of an alternative contrast has no effect on the sums of squares computed for the within subject effect or its error term.

The coefficients for each of the alternative contrasts for a within-subject effect with 4 levels are presented below.

Deviation Contrasts

Deviation contrasts compares a level with the mean of the other three levels. A contrast is not made for the "reference" category, which can be either the first or last level of the within subject effect. The top-left panel shows A4 (reference category last) as the reference category. The top-right panel shows A1 (reference category first) as the reference category. The labels for the ANOVA table are given in the bottom panels. I don't recall a thesis where deviation contrasts would have been helpful.

Reference Category Last	Reference Category First
Contrast Coefficients	Contrast Coefficients

		A					A						
Dependent Variable	Level 1 vs. Mean	Level 2 vs. Mean	Level 3 vs. Mean	Dependent Variable	Level 2 vs. Mean	Level 3 vs. Mean	Level 4 vs. Mean						
A1	.750	-.250	-.250	A1	-.250	-.250	-.250						
A2	-.250	.750	-.250	A2	.750	-.250	-.250						
A3	-.250	-.250	.750	A3	-.250	.750	-.250						
A4	-.250	-.250	-.250	A4	-.250	-.250	.750						
a The contrasts for the within subjects factors are: A: Deviation contrast				a The contrasts for the within subjects factors are: A: Deviation contrast									
Contrast Labels in ANOVA Source Table				Contrast labels in ANOVA Source Table									
Tests of Within-Subjects Contrasts							Tests of Within-Subjects Contrasts						
Source	A	SS	df	MS	F	Sig.	Source	A	SS	df	MS	F	Sig.
A	Level 1 vs. Mean						A	Level 2 vs. Mean					
	Level 2 vs. Mean							Level 3 vs. Mean					
	Level 3 vs. Mean							Level 4 vs. Mean					
Error (A)	Level 1 vs. Mean						Error (A)	Level 2 vs. Mean					
	Level 2 vs. Mean							Level 3 vs. Mean					
	Level 3 vs. Mean							Level 4 vs. Mean					

Simple Contrasts

Simple contrasts compare each cell with a reference cell. The reference cell can be either the last level of the factor (A4 in this example, see the top-left panel) or the first level of the factor (A1 in this example, see the top-right panel). Simple contrasts can be useful if you have a single control group (used as the :reference category) and three treatment groups. This contrast is similar to Dunnett's post-hoc test in that both tests compare a single group with each of the other groups.

Reference Category Last				Reference Category First									
Contrast Coefficients				Contrast Coefficients									
	A				A								
Dependent Variable	Level 1 vs. Level 4	Level 2 vs. Level 4	Level 3 vs. Level 4	Dependent Variable	Level 2 vs. Level 1	Level 3 vs. Level 1	Level 4 vs. Level 1						
A1	1	0	0	A1	-1	-1	-1						
A2	0	1	0	A2	1	0	0						
A3	0	0	1	A3	0	1	0						
A4	-1	-1	-1	A4	0	0	1						
a The contrasts for the within subjects factors are: A: Simple contrast				a The contrasts for the within subjects factors are: A: Simple contrast									
Contrast Labels in ANOVA Source Table				Contrast labels in ANOVA Source Table									
Tests of Within-Subjects Contrasts							Tests of Within-Subjects Contrasts						
Source	A	SS	df	MS	F	Sig.	Source	A	SS	df	MS	F	Sig.
A	Level 1 vs. Level 4						A	Level 2 vs. Level 1					
	Level 2 vs. Level 4							Level 3 vs. Level 1					
	Level 3 vs. Level 4							Level 4 vs. Level 1					
Error (A)	Level 1 vs. Level 4						Error (A)	Level 2 vs. Level 1					
	Level 2 vs. Level 4							Level 3 vs. Level 1					

Level 3 vs. Level 4								Level 4 vs. Level 1						
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Difference Contrasts

Difference contrasts compare: level 1 with level 2, level 3 with the mean of the previous two levels, level 4 with the mean of the previous three levels, and so forth. There is no option to select a "Reference Category." I haven't come across a thesis where difference contrasts would have been helpful.

Contrast Coefficients						
	A					
Dependent Variable	Level 2 vs. Level 1	Level 3 vs. Previous	Level 4 vs. Previous			
A1	-1.000	-.500	-.333			
A2	1.000	-.500	-.333			
A3	.000	1.000	-.333			
A4	.000	.000	1.000			
a The contrasts for the within subjects factors are: A: Difference contrast						
Contrast Labels in ANOVA Source Table						
Tests of Within-Subjects Contrasts						
Source	A	SS	df	MS	F	Sig.
A	Level 2 vs. Level 1					
	Level 3 vs. Previous					
	Level 4 vs. Previous					
Error(A)	Level 2 vs. Level 1					
	Level 3 vs. Previous					
	Level 4 vs. Previous					

Helmert Contrasts

Helmert contrasts compare: the first level of the factor with all later levels, the second level with all later levels, the third level with all later levels, and so forth. There is no option to select a Reference Category. I don't recall a thesis where Helmert contrasts would have been helpful.

Contrast Coefficients			
Dependent Variable	A		
	Level 1 vs. Later	Level 2 vs. Later	Level 3 vs. Level 4
A1	1.000	.000	.000
A2	-.333	1.000	.000
A3	-.333	-.500	1.000
A4	-.333	-.500	-1.000

a The contrasts for the within subjects factors are:
A: Helmert contrast

Contrast Labels in ANOVA Source Table						
Tests of Within-Subjects Contrasts						
Source	A	SS	df	MS	F	Sig.
A	Level 1 vs. Later					
	Level 2 vs. Later					
	Level 3 vs. Level 4					
Error(A)	Level 1 vs. Later					
	Level 2 vs. Later					
	Level 3 vs. Level 4					

Repeated Contrasts

Repeated contrasts compare consecutive pairs of levels, for example level 1 with level 2, level 2 with level 3, level 3 with level 4, and so forth. There is no option to select a "Reference Category." Repeated contrasts can be useful when you have a time-based repeated measures design, as in the study presented in this set of notes.

Contrast Coefficients	

	A		
Dependent Variable	Level 1 vs. Level 2	Level 2 vs. Level 3	Level 3 vs. Level 4
A1	1	0	0
A2	-1	1	0
A3	0	-1	1
A4	0	0	-1

a The contrasts for the within subjects factors are:
A: Repeated contrast

Contrast Labels in ANOVA Source Table

Tests of Within-Subjects Contrasts

Source	A	SS	df	MS	F	Sig.
A	Level 1 vs. Level 2					
	Level 2 vs. Level 3					
	Level 3 vs. Level 4					
Error(A)	Level 1 vs. Level 2					
	Level 2 vs. Level 3					
	Level 3 vs. Level 4					

In this time-based, repeated measures design the repeated contrasts are useful in interpreting the time main effect. The repeated contrasts for the data in this set of notes are shown in Table 13. The summary analysis of variance table with 3 df for the time main effect is the same as that shown in Table 7. The overall effects are the same no matter which set of standard contrasts you select.

Table 13. Tests of Within-Subjects, Repeated, Contrasts

Measure: IES_INT

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
TIME	Level 1 vs. Level 2	4737.633	1	4737.633	62.639	.000	.684
	Level 2 vs. Level 3	16.133	1	16.133	.663	.422	.022
	Level 3 vs. Level 4	50.700	1	50.700	2.662	.114	.084
Error(TIME)	Level 1 vs. Level 2	2193.367	29	75.633			

	Level 2 vs. Level 3	705.867	29	24.340			
	Level 3 vs. Level 4	552.300	29	19.045			

Writing up the results

The significant main effect for measurement time was analyzed by single degree of freedom, "repeated" contrasts. Effect sizes were computed as partial Eta squared values. The contrasts indicate that there was a significant decrease in the IES-Intrusion scores from the pretest ($M = 18.57$, $SD = 8.29$) to the immediate posttest ($M = 6.00$, $SD = 7.31$), $F(1, 29) = 62.64$, $p < .0005$. This treatment effect accounted for 68% of the variability in the IES-Intrusion scores. There was no change from the posttest to the 3-month follow-up ($M = 5.27$, $SD = 6.88$), $F(1, 29) = 0.66$, $p < .442$, nor from the 3-month follow-up to the 15-month follow-up ($M = 3.97$, $SD = 5.93$), $F(1, 29) = 2.66$, $p < .114$.

This way of analyzing the differences between the means ignores the difference that was found between the immediate posttest mean and the 15-month follow-up mean when all pairwise comparisons were made. You would need to decide how important that difference was to your study. In this case the change is small and you might want to decrease the emphasis placed on that difference by just reporting the results of these single degree of freedom contrasts.

top 

9. References

Horowitz, M. J., Wilner, N., & Alvarez, W. (1979). Impact of event scale: A measure of subjective distress. *Psychosomatic Medicine*, *41*, 209-218.

Wilson, S. A., Becker, L. A., & Tinker, R. A. (1995). Eye movement desensitization and reprocessing (EMDR) treatment for psychologically traumatized individuals. *Journal of Consulting and Clinical Psychology*, *63*, 928-937.

Wilson, S. A., Becker, L. A., & Tinker, R. A. (1997). Fifteen-month follow-up of eye movement desensitization and reprocessing (EMDR) treatment for posttraumatic stress disorder and psychological stress. *Journal on Consulting and Clinical Psychology*, *65*, 1047-1056.

top 