

Chapter 12

Repeated Measures One-way ANOVA: Differences Between Multiple Correlated Group Means

Raymond O'Connor, Jr.
Valparaiso University

12.1 Introduction to Repeated Measures One-way ANOVA

As in Chapter 11, the topic of this chapter is also One-Way ANOVA. Thus, much of what you learned in Chapter 11 will be applicable to this chapter. The major difference is that whereas we introduced **between groups** independent samples ANOVA in Chapter 11, in the present chapter we will discuss **within group** correlated samples ANOVA.

You may recall that one instance of a within groups design involves repeated measures on the same participants. The **before-after** design introduced in Chapter 10 is an example of this type of design. Recall that EZ employees were measured on performance, social skills and task skills before they attended a leadership workshop (**soc, task, and perform**), then they were re-measured on these variables immediately after the workshop (**soc2, task2, and perform2**).

In Chapter 10, we introduced the **paired-samples t-test** as the appropriate analysis for comparing scores before and immediately after the workshop. However, you may recall from Chapter 4 that we also included a **third measurement** of these variables three months after the workshop. This *delayed post test* is not uncommon with program evaluation. That is, researchers are not only interested in immediate gains after participation in the workshop, they want to know whether or not the change is long-lasting.

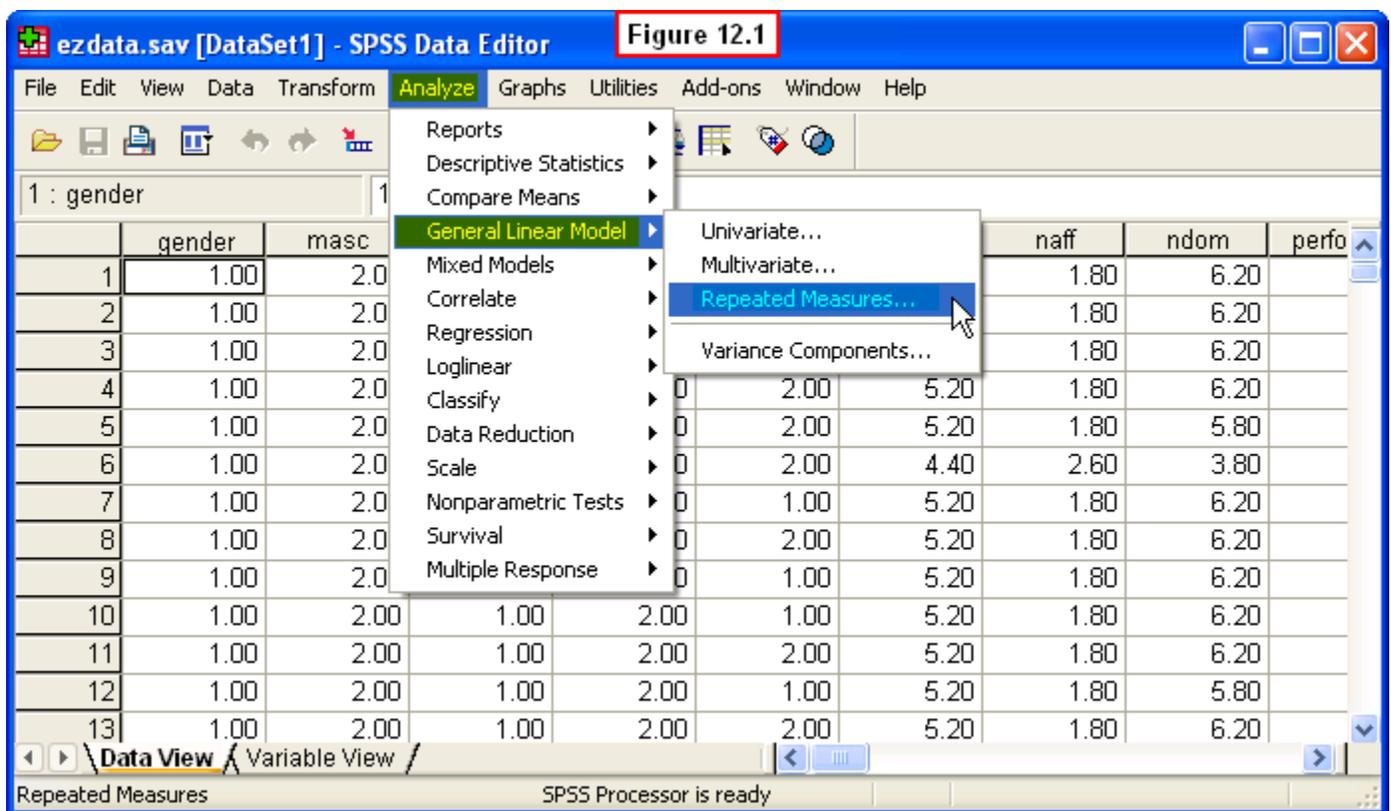
Thus, while we saw in Chapter 10 that there was a significant improvement in scores after the workshop, the focus in the present chapter is on whether or not that change was enduring, i.e., will the increases be maintained over three months? While the optimal situation would be to see the same level of improvement at three months, there is typically some loss over time. So many program evaluators consider it a success if the three-month post test shows that scores are still significantly higher than at pre-test, even if they may be lower than observed immediately after the workshop. The analyses introduced in this chapter will answer this question.

Thus, this design bears similarities to the two-group repeated measures t-test design in Chapter 10, but also shares characteristics of independent samples one-way ANOVA in that there are now three groups to be compared. So the appropriate analysis of this data is a **repeated measures one-way ANOVA**.

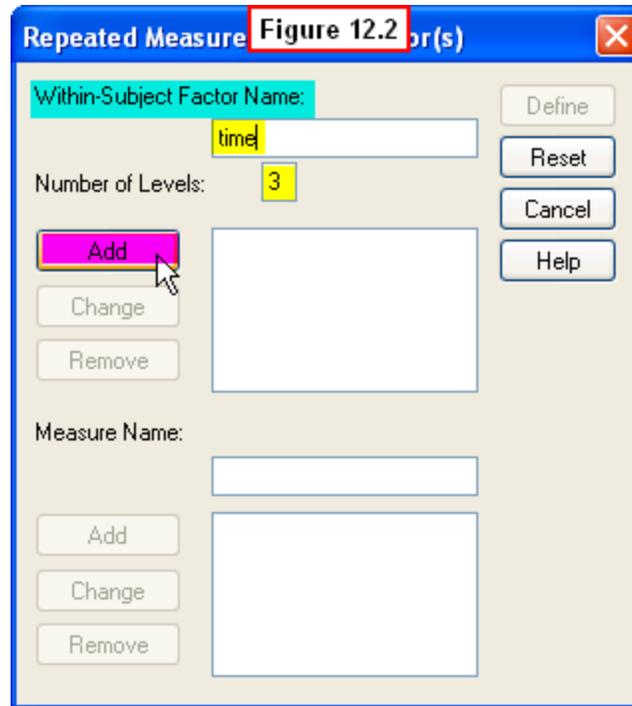
For the example of this design and analysis in this chapter, we will determine whether or not there are significant differences in the leadership performance scores at the three times of measurement (before, immediately after, and three months after the workshop). These variables are **perform**, **perform2**, and **perform3** in the ezdata file. For the chapter exercise, we will ask you to examine differences in the leader social skills scores across the three times (**soc**, **soc2**, and **soc3**).

12.2 Running Repeated Measures ANOVA

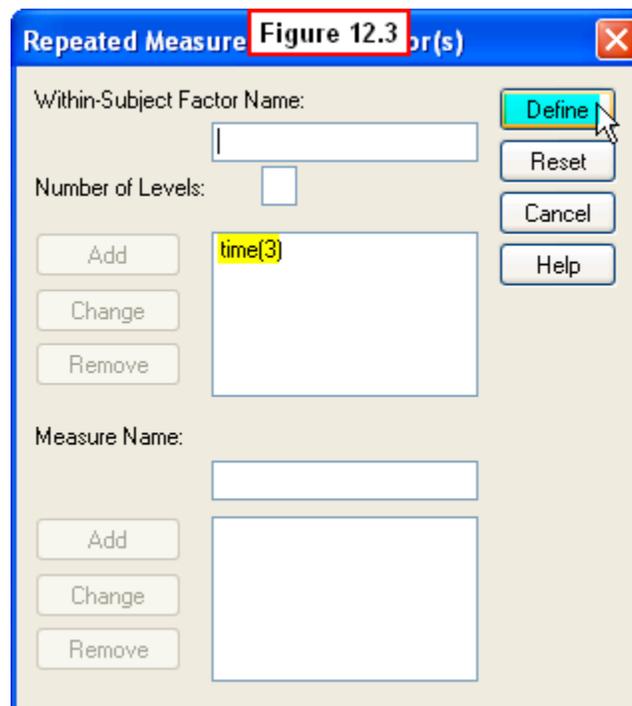
Open your **ezdata.sav** file and select **Analyze, General Linear Model, Repeated Measures...** from the menu (Figure 12.1).



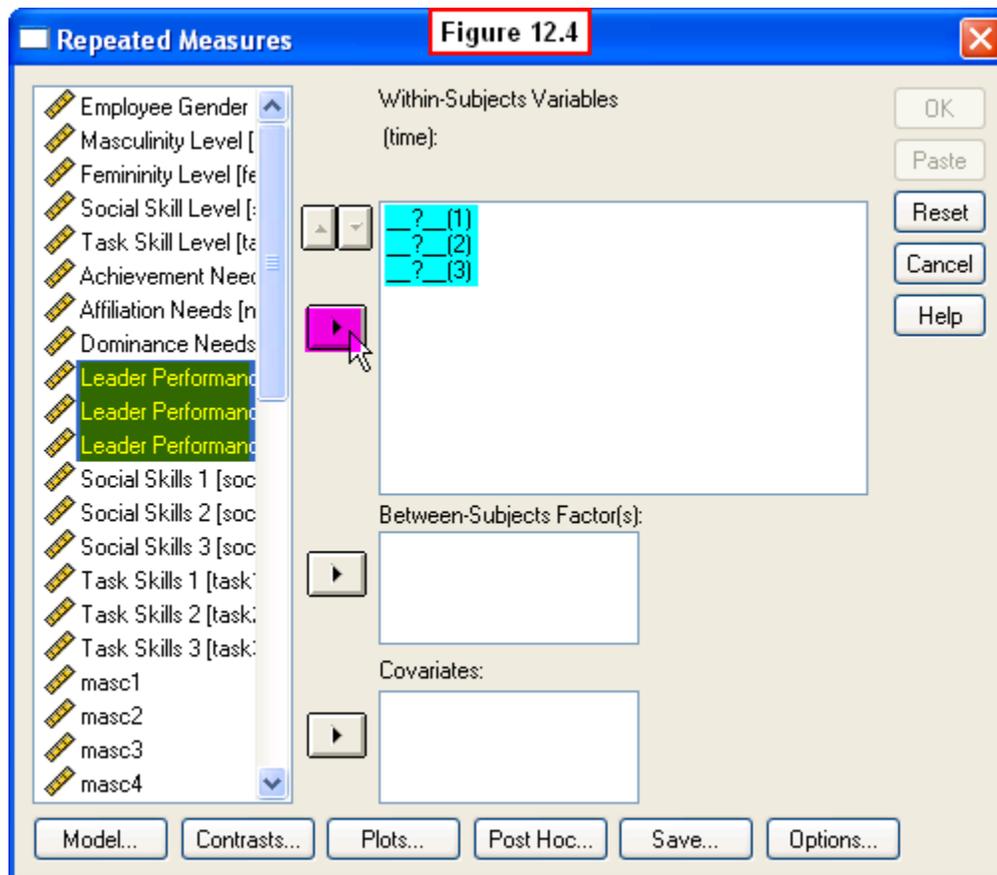
A **Repeated Measures Define Factor(s)** dialog window will appear (Figure 12.2). We need to define the **Within-Subject Factor Name** (i.e., the repeated measures variable, **perform**, **perform 2**, **perform3**), and enter the number of levels of this factor. To do this, type **time** in the box below **Within-Subject Factor Name**, and enter a **3** in the box next to **Number of levels**. Then click the **Add** button.



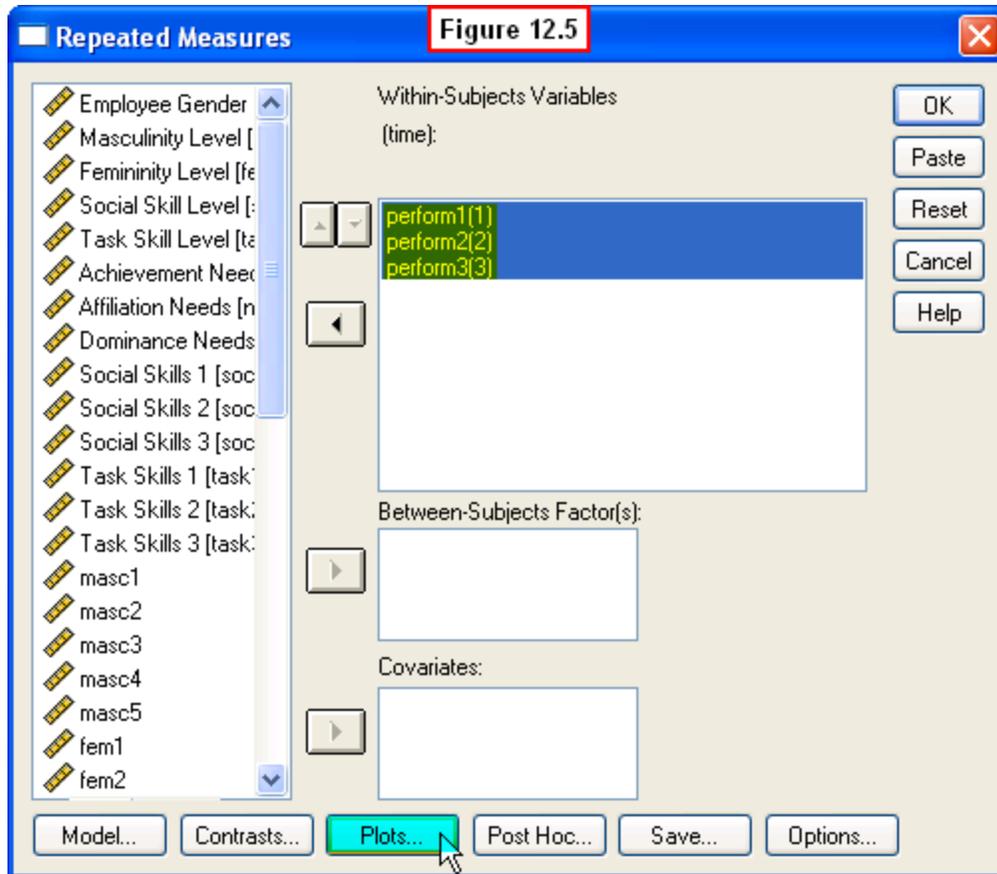
After you have done this, your window should look like that shown in Figure 12.3. Note that SPSS will now name this variable **time**, and it shows that there are three levels: **time (3)**. Now SPSS needs us to define what the three levels of time are. To do this, click the **Define** button in the upper right corner.



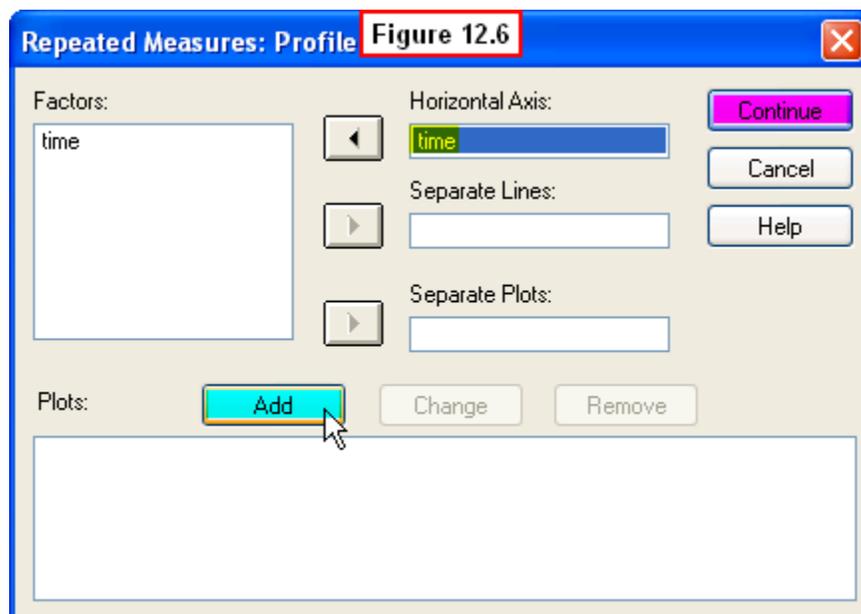
A **Repeated Measures** dialog window will appear (Figure 12.4). To specify the levels, scroll down the variable list on the left, then highlight the three **Leadership Performance** variables. Click the right-arrow button in the middle, and these variables will be moved to the appropriate places in the **Winthin-Subjects Variables (time)** variable definition box.



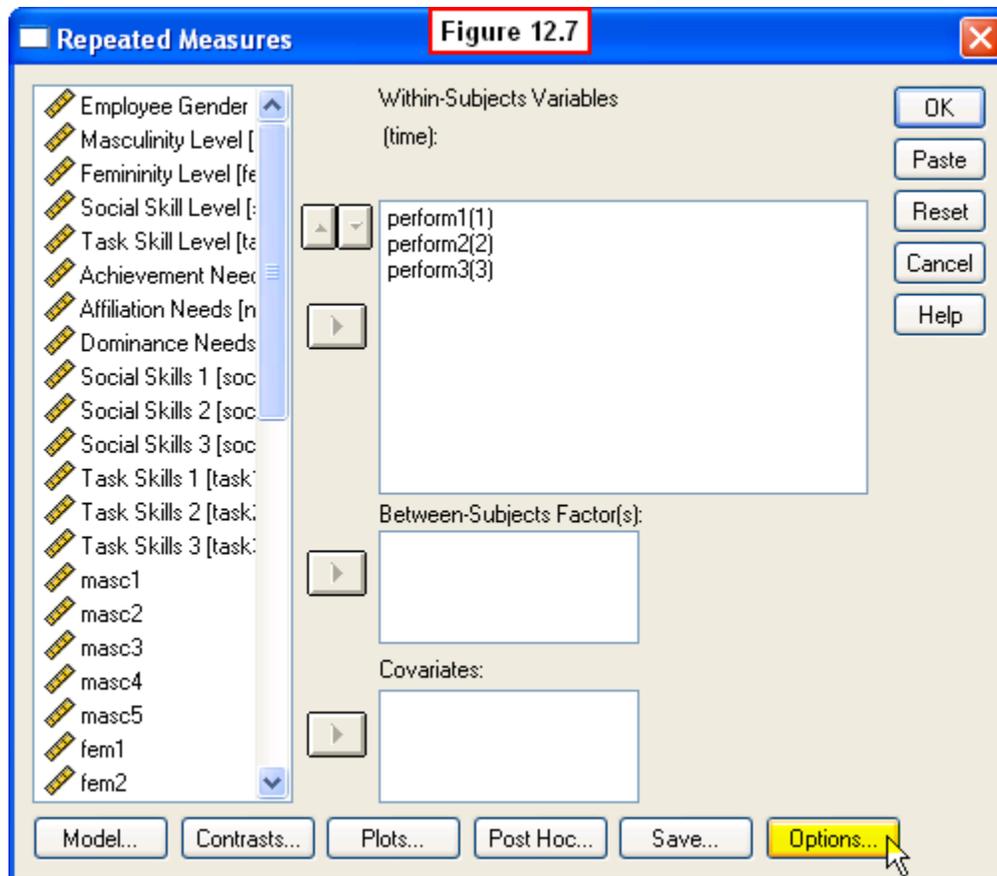
Your window should now look like the one shown in Figure 12.5. As we saw in Chapter 11, it is helpful to see a graph of the three performance score means, so to generate this, click the **Plots...** button at the bottom of this window.



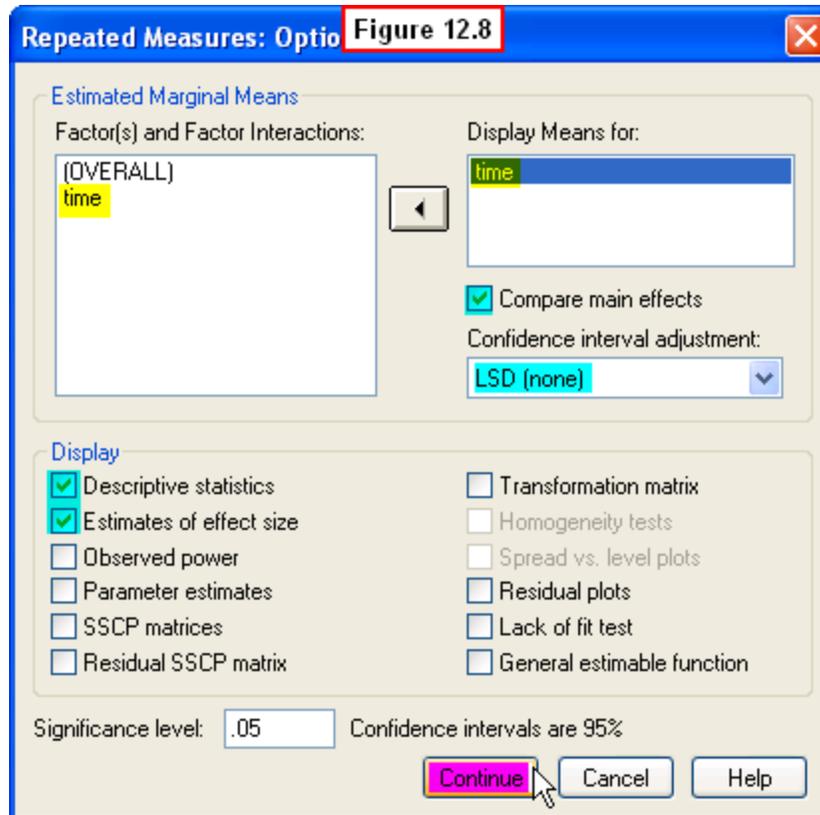
A **Repeated Measures: Profile Plots** dialog window will appear (Figure 12.6). Select the **time** variable on the left, then click the right-arrow button in the middle to move this to the **Horizontal Axis** box on the right. Then click the **Add** button at the bottom of this dialog window.



This will cause SPSS to create a figure displaying the mean performance scores on the Y-axis (vertical) and the three times of measurement on the X-axis (horizontal). Click the **Continue** button in the upper-right corner, and this will close this window and return you to the main **Repeated Measures** dialog window (Figure 12.7).



So far we have instructed SPSS to compute the repeated measures one-way ANOVA. Recall from Chapter 11 that the overall ANOVA results just tell us whether or not there is a significant difference *somewhere* between the three means. In order to determine which means are significantly different, we need to instruct SPSS to compute multiple comparisons of means following the ANOVA. To do this, click the **Options...** button in the lower-right corner. A **Repeated Measures: Options** dialog window will appear (Figure 12.8).



First highlight the **time** variable on the left and move it to the **Display Means for:** box on the right. Then check the **Compare main effects** box and select **LSD (none)** from the drop down menu in this box. Recall from Chapter 11 that the LSD test is a procedure for conducting multiple comparisons among the three mean performance scores.

Next, check the **Descriptive statistics** box to generate other descriptive statistics on these variables, and check the **Estimates of effect size** box to generate partial-eta square values to assess how large the differences between the means is (see Chapter 11). Last, click the **Continue** button at the bottom of this window.

This window will close and return you to the main **Repeated Measures** dialog window. To run our analysis, click the **OK** button in the upper right corner of this window. The results of this analysis will then be displayed in an output viewer window.

12.3 Interpreting the Output

The first table of the output simply repeats the definition of our within-groups factor, **time**. Next, the **Descriptive Statistics** table (Figure 12.9) presents the three performance means, their standard deviations and sample sizes. The most relevant statistics for our purposes are the three means reflecting leader performance scores.

Examination of these means suggests that the average leadership performance was lowest at time 1 (before the workshop), highest at time 2 (immediately after the workshop), and in between these two at time 3 (the 3-month follow-up measure). But before we can interpret these means, we must first examine the results of the *F*-test and the multiple comparisons of means. Also, skip over the next table on the output, **Multivariate Tests**, since it is beyond the scope of this chapter.

Figure 12.9

Descriptive Statistics

	Mean	Std. Deviation	N
Leader Performance 1	5.9211	2.36772	228
Leader Performance 2	6.9561	1.75047	228
Leader Performance 3	6.3816	1.80284	228

Before we examine the ANOVA results, we must first examine the **Mauchly's Test of Sphericity** (Figure 12.10). Complete discussion of this test is also beyond our scope, but interpretation of this statistic is similar to that of the Levine's Test for Equality of Variance (see Chapter 9, section 9.3a). Since the **Sig** probability (.000) is less than .05, we cannot assume that the variances between the three sets of scores are equal.

Figure 12.10

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	.250	313.637	2	.000	.571	.572	.500

As we saw in Chapter 9, this means that a statistical correction needs to be made due to the lack of homogeneity (equality) of variance. SPSS provides this adjustment in the next table of the output (Figure 12.11). This table, **Tests of Within-Subjects Effects**, presents the ANOVA results. Note that there are four *F*-values shown in the first section of this table. The first row presents the results if sphericity could be assumed. The latter three values are different ways of statistically correcting for the lack of sphericity. The most conservative of the three is the **Lower-bound** test (which is highlighted in this table). In actuality, there was not much of an adjustment in that all four *F*-values are the same.

Figure 12.11

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
time	Sphericity Assumed	122.635	2	61.317	58.479	.000	.205
	Greenhouse-Geisser	122.635	1.143	107.328	58.479	.000	.205
	Huynh-Feldt	122.635	1.145	107.142	58.479	.000	.205
	Lower-bound	122.635	1.000	122.635	58.479	.000	.205
Error(time)	Sphericity Assumed	476.032	454	1.049		↑	↑
	Greenhouse-Geisser	476.032	259.374	1.835			
	Huynh-Feldt	476.032	259.824	1.832			
	Lower-bound	476.032	227.000	2.097			

As we have seen in previous discussions of null hypothesis testing, since the **Sig.** level is less than .05, then we can reject the null hypothesis of no difference in performance between the three times of measurement in favor of the alternative hypothesis that there is a significant difference somewhere between the three performance means. The partial eta squared value (.205) indicates that this is a relatively small effect, even though it is significant.

Skip over the next table, **Tests of Within-Subject Contrasts**, since this analysis is not relevant here. Also skip over the next table of the output, **Tests of Between-Subjects Effects**, since there is no between-groups factor in this design. The next table, **Estimates**, again presents the three means, their standard errors and upper/lower bounds.

We are interested in comparing these means, so the next table, **Pairwise Comparisons**, is the one we need to examine. The multiple comparisons of the three means are presented in this table (Figure 12.12).

Figure 12.12

Pairwise Comparisons

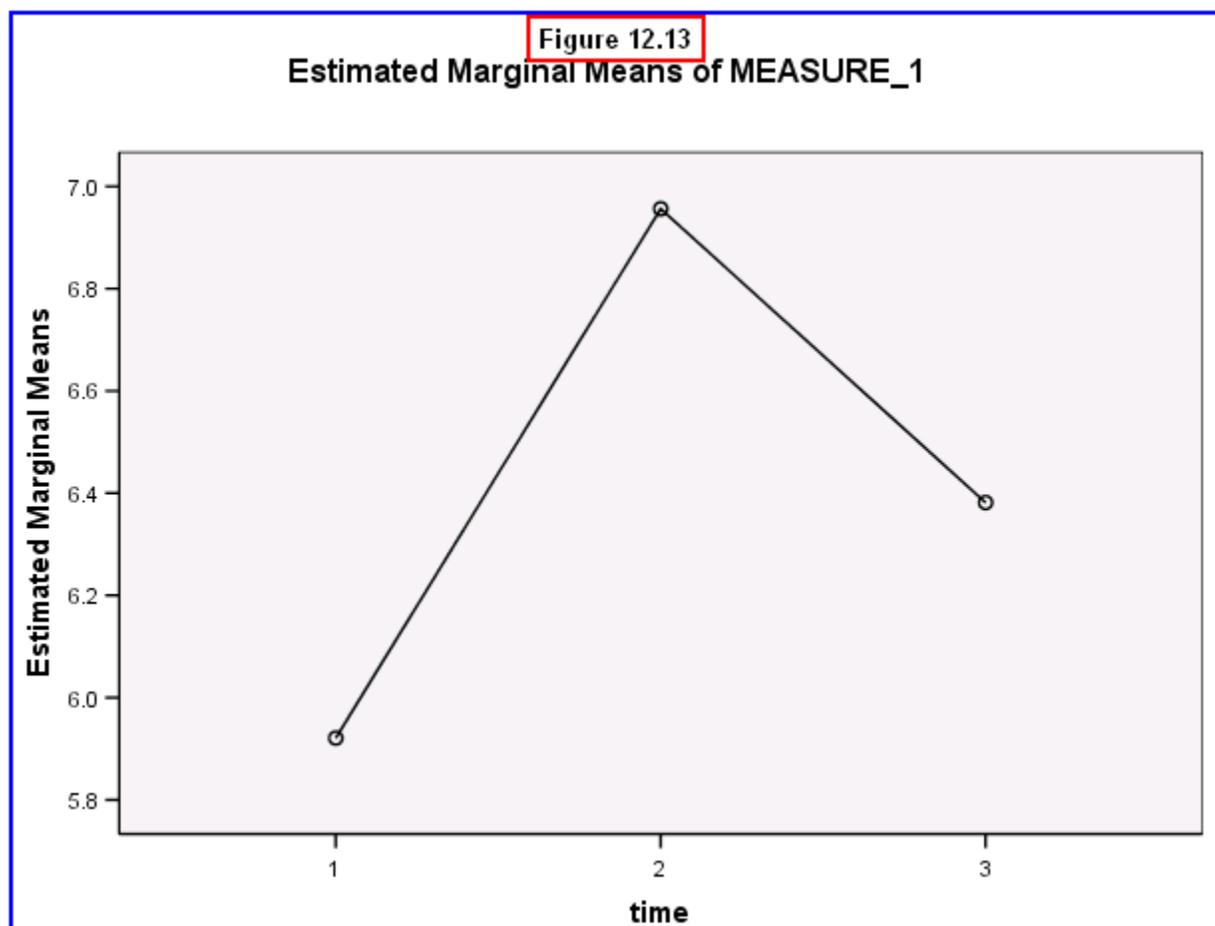
Measure: MEASURE_1

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
1	2	-1.035*	.126	.000	-1.282	-.788
	3	-.461*	.099	.000	-.656	-.265
2	1	1.035*	.126	.000	.788	1.282
	3	.575*	.045	.000	.486	.663
3	1	.461*	.099	.000	.265	.656
	2	-.575*	.045	.000	-.663	-.486

The first section of this table presents the comparisons of the mean performance scores at time 1 to those at time 2 and time 3. The second column shows the actual difference between each pair of means being compared. Our primary interest is in the **Sig.** column, because this indicates whether or not any given pair of means is significantly different. Since $p < .05$ in this column for both comparisons, we can conclude that performance at time 1 was significantly different from that at both time 2 and time 3. Note that this is also indicated by an asterisk placed next to the Mean Difference value.

Since we have already made two comparisons, all that remains is the comparison of the means at time 2 and time 3. This is shown in the second section of this table. The difference between time 2 and time 3 (i.e., .575) is statistically significant, since $p < .05$ for this comparison in the **Sig.** column. In sum, these comparisons reveal that all three means are significantly different from each other. The remainder of this table is redundant, since we have already discussed all three comparisons.

Scroll to the end of the output file to see the graph of these three means. The pattern of differences just described can be seen clearly in the plot of the three means shown in Figure 12.13.



This graph illustrates that performance increased significantly immediately after the workshop, but that performance was significantly lower three months later than immediately after the workshop. However, the performance at three months was still significantly higher than it was before the workshop. Below is an example of how these results would be written in APA format:

A repeated measures one-way ANOVA revealed that there were significant differences in leadership performance between the three times of measurement, $F(1,227) = 58.48, p < .001$, though this was a relatively small effect size (eta-squared = .21). LSD comparisons revealed that all three means were significantly different from each other. Mean leadership performance was significantly higher immediately after the leadership training workshop ($M = 6.96$) than before the workshop ($M = 5.92$). The mean performance score three months later was significantly lower ($M = 6.38$) than that immediately after the workshop, but significantly higher than the mean performance before the workshop.

The repeated measures one way ANOVA procedure we performed yielded interesting results that you would want to report to upper-level management at EZ Manufacturing. In particular, there is evidence that the leadership training program did significantly increase performance. While the largest increase was observed immediately after the workshop, and there was a significant decline after three month, the performance at three months was still significantly higher than before the workshop. This indicates that the workshop was effective at producing improved leadership, and that this increase was relatively enduring.

12.4 Chapter Review Video

[Review Me!](#)

12.5 Try It! Exercises

1. Using Repeated Measures One-Way ANOVA: Assessing Changes in Leader Social Skills

Use the procedures described in Section 12.2 to compute a repeated-measures one-way ANOVA to test for differences in leader social skills before, immediately after, and three months after EZ employees attended the leadership training workshop. This variable has three levels (**soc1**, **soc2** and **soc3**). Define this within-groups variable as **time**. Generate a plot of the means and request LSD comparisons by choosing the **Compare Main effects** option.

When you have generated your output file:

- **Print** your output file to submit to your instructor.
- Answer the following questions:
 - What is the value of the F -statistic?
 - What is the significance level?
 - Should you reject or fail to reject the null hypothesis of no differences in social skills across the three time periods?
 - Write a statement about the F -value as shown in Section 12.4
 - Based on the results of the LSD Multiple Comparisons tests (pairwise comparisons), which means are significantly different?
 - For those means that are significantly different, state what the differences are (i.e., state which social skills means are higher or lower than the others).
- **Write an interpretation of the results in APA format** to submit to your instructor. Follow the example in Section 12.4.