
15

REPEATED MEASURES PROCEDURES: T_{DEP} AND ANOVA_{WS}

I have mentioned several times in this book that the researcher needs to consider a statistical procedure carefully before using it. The greatest power and confidence relies on using the correct statistical procedure with the available data according for the research design.

In this chapter, I am returning to two statistical procedures that statisticians have adapted for use with “repeated measures” data. I discussed the independent T test and the general analysis of variance (ANOVA) procedure in earlier chapters, both of which were designed for independent samples. While there are other such procedures, the “dependent T test” and the “within-subject ANOVA” represent important procedures that researchers can use in experimental and post facto designs in which the data are “linked” in some way.

Recall that I made a distinction between statistical procedures that used independent samples and those that used dependent samples. In Chapter 8, I identified dependent sample designs as those in which one group is somehow linked structurally to the other. Thus, in experimental designs, a group measured twice (e.g., pretest and posttest) would need to be treated differently than other group measures because the *same group of subjects would be measured twice*.

If a researcher used a T test to detect a difference between two dependent samples in an experiment (i.e., pre–post, Time1–Time2, matched samples), they would need to use a special T test that factored out the relatedness, so the T test would reveal the unique differences that remained between the two samples. The same thing would be true for ANOVA procedures that measured the same group twice.

Using Statistics in the Social and Health Sciences with SPSS® and Excel®, First Edition.

Martin Lee Abbott.

© 2017 John Wiley & Sons, Inc. Published 2017 by John Wiley & Sons, Inc.

As I noted in Chapter 14 in which I discussed the dependent sample chi-square (McNemar's test), these kinds of measures have several names:

- Repeated measures tests
- Dependent sample tests
- Paired sample tests
- Within-subject tests

INDEPENDENT AND DEPENDENT SAMPLES IN RESEARCH DESIGNS

Independent and dependent sample data can be used in both experimental and post facto designs. Dependent samples are most commonly encountered in experimental or quasi-experimental designs in which a sample group is given a pretest and then administered a posttest after some intervention. It is used in some post facto research to measure a group on some attitude measure at different times.

Table 15.1 shows several possibilities for independent and dependent samples in research and post facto designs. This table highlights the dependent T test since there are only two groups. A similar table could be created for three or more groups that would emphasize the within-subjects ANOVA.

As I noted in Chapter 8, independent samples mean that choosing subjects for one group has nothing to do with choosing subjects for the other group. Thus, if I randomly select Bob and assign him randomly to group 1 (the experimental group

TABLE 15.1 Independent and Dependent Sample in Experimental and Post Facto Designs

	Experimental Designs	Post Facto Designs
Two Independent Samples (Between Subjects)	T_{ind} comparing Post test differences of two groups of <i>different subjects</i> after an intervention <i>Example: Do the 'sensitivity to violence' (STV) measures differ between two randomly chosen groups after one group watches a violent movie?</i>	T_{ind} comparing existing data on two groups of <i>different members</i> <i>Example: Do STV measures differ between Trauma workers and GP's?</i>
Two Dependent Samples (Within Subjects)	T_{dep} comparing Pre—Post differences or Time 1—Time2 differences of the <i>same group of subjects</i> after an intervention <i>Example: Is the post test STV score higher than the pre test PFV scores of one group that watches a violent movie?</i>	T_{dep} test comparing <i>matched group</i> differences on an existing measure or outcome <i>Example: Do the STV scores differ between two groups of doctors (one in the urban core and one in the suburbs) matched on time as a physician?</i>

that watches a violent movie), it has nothing to do with the fact that I choose Sally and assign her randomly to group 2 (the group that does not see the violent movie). The random selection of these group members is a very important assumption because it assures the researcher that there are no built-in linkages between the groups’ posttest scores. The power of randomization will result in the comparability of two groups chosen in this way.

Dependent samples consist of groups of subjects that have some structured linkage, like using the same people twice in a study. For example, I might use pretest STV scores from both Bob and Sally and *compare these scores with their own posttest STV scores*. Using dependent samples affects the ability of the randomness process to create comparable samples; in such cases, the researcher is assessing *individual* change (before-to-after measures) in the context of the experiment that is assessing *group* change.

Dependent samples also include *matched samples*, a situation in which the researcher purposely assigns people to be in two groups on the basis of some characteristic(s) rather than choosing and assigning subjects randomly to two groups. For example, you might be concerned about gender for a given study and purposely assign equal numbers of men and women to two groups. In this case, the randomness criterion is violated despite the attempt of the researcher to ensure that study groups have matching characteristics on a crucial study influence. This is typically the case where the researcher cannot control the conditions for random selection and assignment.

USING DIFFERENT T TESTS

When researchers use dependent samples, they need to use an independent T test that has been “adjusted” for the influence of the relatedness of the samples. This adjusted test is called a *dependent sample T test*. Other names for this test are repeated measures T test, within-subject T test, and paired sample T test. Both Excel and SPSS® refer to these as paired sample T tests.

There are two ways to calculate the dependent T test (T_{dep}). I will introduce both approaches in this chapter, but I will focus primarily on interpreting the SPSS and Excel procedures rather than the manual calculations.

THE DEPENDENT T TEST CALCULATION: THE “LONG” FORMULA

The long formula for T_{dep} looks intimidating, but it makes conceptual sense. It is essentially the same as the formula for T_{ind} except that you must adjust the estimated standard error of difference s_D to account for the relatedness in the samples.

As you will recall from Chapter 8, these are the T_{ind} formulas:

$$t = \frac{(M_1 - M_2) - \mu_{M_1 - M_2}}{s_D}$$

$$s_D = \sqrt{s_{m_1}^2 + s_{m_2}^2}$$

The formulas for T_{dep} are the same except for the estimated standard error of difference using the adjusted formula in the following (s_{D_r}):

$$t_{\text{dep}} = \frac{(M_1 - M_2) - \mu_{M_1 - M_2}}{s_{D_r}}$$

The T_{dep} formula thus becomes

$$t_{\text{dep}} = \frac{(M_1 - M_2)}{s_{D_r}}$$

Here is the formula for the adjusted standard error of difference:

$$s_{D_r} = \sqrt{s_{M_1}^2 + s_{M_2}^2 - [2r(s_{M_1})(s_{M_2})]}$$

As you can see, the only difference is that there is a “second component” included that is subtracted out of the s_D in the T_{ind} formula. I shaded this portion in the formula. Recall that I said that dependent samples are related (correlated) to one another and that the dependent T test must subtract out this relatedness.

If you look at the T_{dep} formula, the shaded portion that measures this relatedness is subtracted out of the formula. The shaded portion includes “ r ,” the Pearson’s correlation coefficient. If the two sample groups are related (correlated), this r value will “modify” the measures that are used to calculate the T_{dep} ratio. However, if you have two independent samples, they will not be related, r will equal 0, and the shaded portion will therefore also equal 0. When this is the case, the s_{D_r} will be equivalent to s_D .

EXAMPLE: THE LONG FORMULA

I will use an example to show these calculations briefly. Recall the formulas I used to create the T_{ind} which I will use to assist in creating the T_{dep} values:

$$s_x = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N - 1}} \text{ or } S_x = \sqrt{\frac{SS}{df}}$$

$$s_D = \sqrt{s_{m_1}^2 + s_{m_2}^2} \quad s_{m_1} = \frac{s_1}{\sqrt{n_1}} \quad s_{m_2} = \frac{s_2}{\sqrt{n_2}}$$

The data for this exercise are fictitious. The values represent the amount of time it takes (number of minutes) for lab technicians to complete a routine tissue test, from receipt of the order to logging the results. Time1 (in the two shaded columns) represents the initial information gathered from the 45 lab technicians (from different sections of a large hospital). Next, the lab technicians are given instruction in new procedures for performing the lab test. Time2 (in the nonshaded columns) represents the testing time from the same technicians as in Time1. Since the same lab technicians

EXAMPLE: THE LONG FORMULA

493

were assessed at two different times, you would use the T_{dep} to see if the intervention (instruction in the new procedure) was effective.

Table 15.2 shows the data in columns. Because of space limitations, I created two columns each for Time1 and Time2. Use the formulas I included earlier to calculate the paired T ratio.

Table 15.3 shows the relevant calculations for this problem. Confirm the summary calculations and then use them to conduct a T_{dep} test.

Results

$$\begin{aligned}
 t_{\text{dep}} &= \frac{(M_1 - M_2)}{s_{D_r}} \\
 &= \frac{(40.51 - 36.76)}{0.888} \\
 &= 4.223
 \end{aligned}$$

TABLE 15.2 The Study Data

Time1	Time2	Time1	Time2
11	6.0	41	35.0
4	2.0	62	47.0
11	11.0	29	29.0
18	9.0	55	50.0
30	23.0	46	49.0
7	12.0	39	38.0
12	10.0	41	40.0
42	42.0	45	46.0
32	20.0	49	58.0
17	5.0	41	35.0
24	24.0	54	56.0
46	37.0	50	53.0
26	28.0	52	58.0
28	31.0	50	49.0
44	39.0	56	57.0
44	30.0	62	60.0
38	38.0	57	39.0
45	39.0	54	48.0
35	34.0	79	70.0
45	36.0	64	62.0
39	31.0	47	41.0
40	37.0	64	50.0
48	40.0		

TABLE 15.3 The Calculated Elements of the Study Data

	Time1	Time2
Means	40.51	36.76
SS	12291.244	12102.311
s_x	16.714	16.585
s_m	2.492	2.472
s_D	3.51	
R	0.936	
s_{D_r}	0.888	

The paired T ratio for this study is 4.223. When you compare this to the tabled value of T to determine whether it is a significant ratio, use $df = N - 1$, where N is the *number of pairs of data*. Therefore, $T_{\text{dep}0.05,44} = 2.021$. Therefore, since the calculated $T_{\text{dep}} = 4.223$, you can reject the null hypothesis ($p < 0.05$). The difference between Time1 and Time2 is statistically different. The lab technicians took less time conducting the tissue test. Time2 data suggest lower times for each lab technician.

Effect Size

You can use the same effect size formula for the T_{dep} as you did for the single sample T test (in Chapter 7). The difference is that N is the number of pairs of data. Using 0.2, 0.5, and 0.8 as the criteria for small, medium, and large, as you did in Chapter 7, you can judge the calculated effect size for this finding ($d = 0.63$) as a medium effect:

$$\begin{aligned} d &= \frac{T_{\text{dep}}}{\sqrt{N}} \\ &= \frac{4.223}{\sqrt{45}} \\ &= 0.63 \end{aligned}$$

THE DEPENDENT T TEST CALCULATION: THE “DIFFERENCE” FORMULA

There is another way of manually calculating paired data that is shorter and follows a procedure you have already learned. As you recall from Chapter 7, I discussed the process for calculating the single sample T ratio. The formulas I used were

$$t = \frac{M - \mu}{s_m} \quad s_m = \frac{s_x}{\sqrt{N}}$$

You can use this method to calculate a T ratio for paired data by subtracting values in one set of scores from paired values in the other set of scores to yield a single set of “difference” scores. Once you have the difference scores, you can use the formulas earlier to calculate the result. You can see how I included the “difference” scores into the single sample T formula in the following to use it with our difference change scores:

$$t = \frac{M - \mu}{s_m} \quad \text{becomes} \quad T_{\text{dep}} = \frac{M_{\text{diff}} - 0}{s_{m-\text{diff}}}$$

One matter of note in the altered formula that uses the difference scores is the “0” in the place occupied by the population mean in the single sample T formula. This simply refers to the fact that the population of all “differences between related group scores” will be 0 since some differences will be negative and some will be positive. This would create a 0 sum. Therefore, the formula compares the *sample* mean difference score to the *population of difference scores*, the latter of which is 0:

$$T_{\text{dep}} = \frac{M_{\text{diff}} - 0}{s_{m\text{-diff}}}$$

$$SS_{\text{diff}} = \sum D^2 - \frac{(\sum D)^2}{N}$$

$$s_{\text{diff}} = \sqrt{\frac{SS_{\text{diff}}}{N-1}}$$

$$s_{m\text{-diff}} = \frac{s_{\text{diff}}}{\sqrt{N}}$$

The bottom line is that you can use the same formula you used with the single sample T test with only slight variations. In this way, you do not have to calculate the Pearson's r value as you did in the “long” formula. By creating difference scores, you “subtract out the relatedness.”

The two approaches (long formula and difference formula) yield the same T_{dep} value. With larger data sets, you will likely use SPSS and Excel to make the calculations, so you do not need to be concerned at this point with the different formulas.

Here is how the difference formulas and procedure would work using the data in Table 15.2. Table 15.4 shows the summary calculations which you can use to conduct the T_{dep} test. Remember that the df for this calculation is $N - 1$ (pairs).

The T_{dep} Ratio from the Difference Method

When you use this method, use the modified formula in the following to calculate the T_{dep} :

$$T_{\text{dep}} = \frac{M_{\text{diff}} - 0}{s_{m\text{-diff}}}$$

$$= \frac{3.756 - 0}{0.887}$$

$$= 4.234$$

TABLE 15.4 The Difference Procedure for Calculating T_{dep}

Time1	Time2
Mean	3.756
$\sum D$	169
$\sum D^2$	2193
SS_{diff}	1558.31
s_{diff}	5.95
$s_{m\text{-diff}}$	0.887

TABLE 15.5 The T_{ind} Comparison with T_{dep}

$T_{\text{ind}} = 1.07$	$T_{\text{dep}} = 4.234$
$T_{.05,88} = 2.00$	$T_{.05,44} = 2.021$
Decision: Do not reject H_0	Reject H_0 , $p < .05$

If you compare the T_{dep} value from the difference method, you will find that it is equivalent to the T_{dep} calculated from the long method, with some rounding discrepancy.

T_{DEP} AND POWER

As I have stated several times, *the maximum efficiency of statistical procedures is reached when the researcher uses the appropriate statistical measure with the study data.* To see how this works in the example study earlier, conduct T_{ind} with the data and compare the results to the T_{dep} result. Table 15.5 shows the comparison. As you can see, if you had not used the appropriate formula (T_{dep}), you would not have been able to reject the null hypothesis. However, since you did use T_{dep} , you rejected the null hypothesis and concluded that the training for the tissue test procedure was effective.

CONDUCTING THE T_{DEP} ANALYSIS USING SPSS®

The T_{dep} analysis with SPSS is straightforward. The T_{dep} analysis is accessed by using the “Analyze – Compare Means – Paired-Sample T Test” menu. This will create the window shown in Figure 15.1.

As you can see in Figure 15.1, I simply used the arrow button to indicate Time1 as Variable1 and Time2 as Variable2 for the analysis. This is all that is necessary. The



Figure 15.1 The SPSS® T_{dep} specification window.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Time1	40.51	45	16.714	2.492
	Time2	36.76	45	16.585	2.472

Figure 15.2 The SPSS® descriptive output for T_{dep} .

default formulas and specifications will produce the analyses needed. When I choose “OK,” SPSS produces a series of output tables that I will use to complete the T_{dep} analysis. I show these output tables in Figures 15.2–15.4.

Figure 15.2 shows the descriptive findings from the data. As you can see the s_m values (2.492 and 2.472) are produced and can be used with the long formula to check manual calculations.

Figure 15.3 shows the correlation analysis between Time1 and Time2 which is necessary for the long formula calculation. In any case, it is important to note even if you choose to calculate the T_{dep} manually with the difference method.

Figure 15.4 shows the T_{dep} test findings. You can compare these to the manual calculations earlier. Note that the s_{D_r} value (0.888 in the manual calculations) is shown as “Std. Error Mean.” The T_{dep} value (4.233) is shown with the calculated significance ($p = 0.000$). The mean shown in the first column of data is the “difference in means” ($40.51 - 36.76 = 3.75$) from Time1 and Time2 that is used in the numerator of the T_{dep} formula. The “95% confidence interval of the difference” values (1.968–5.543) indicates the confidence brackets that will contain the true population difference (i.e., Time1–Time2 difference) 95% of the time. Note that $df = 44$, indicating “pairs – 1.”

Paired Samples Correlations			
		N	Sig.
Pair 1	Time1 & Time2	45	.936

Figure 15.3 The SPSS® correlation output for T_{dep} .

Paired Samples Test									
		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Time1 - Time2	3.756	5.951	.887	1.968	5.543	4.233	.44	.000

Figure 15.4 The SPSS® T_{dep} test summary.

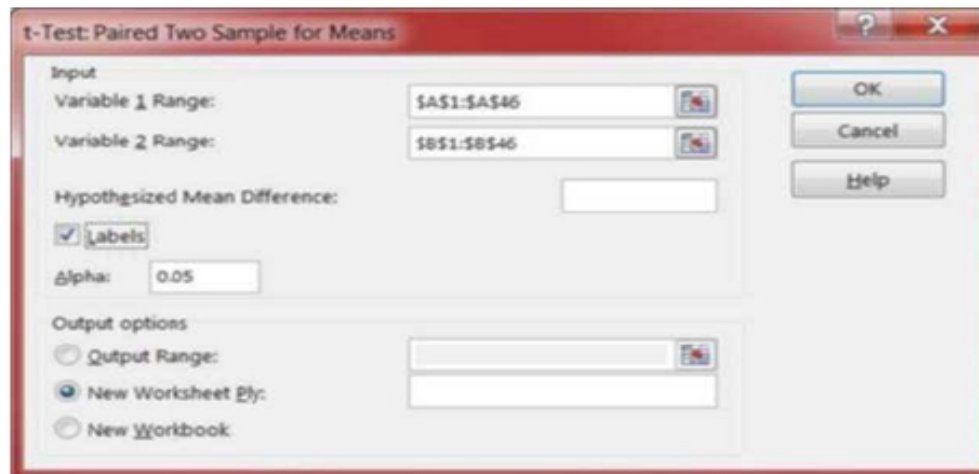


Figure 15.5 The Excel T_{dep} specification window.

CONDUCTING THE T_{DEP} ANALYSIS USING EXCEL

The Excel procedure is created through the Data – Data Analysis – “t-Test: Paired Two Sample for Means” menu. When you make this choice, the screen in Figure 15.5 appears. As you can see, I specified the location of the data in the two “Variable Range” windows, and I indicated that the data included the variable label by checking the “Labels” box.

When you make this selection, Excel returns the findings shown in Figure 15.6. As you can see, the T_{dep} value of 4.23 is shown (shaded cell), and the correlation (0.94) is shown. The values below the T_{dep} figure show the results of the hypothesis test(s) with their respective exclusion values. For example, the cell that includes “ $P(T \leq t)$ one tail” indicates that T_{dep} is significant at “0.00” (i.e., beyond $p < 0.05$) for the one-tailed T test which sets the exclusion value at 1.68. The last two cells confirm the significance level (“0.00” or beyond $p < 0.05$) with a two-tailed test in which the exclusion value = 2.02.

WITHIN-SUBJECT ANOVA (ANOVA_{WS})

I now consider another extension of ANOVA. This time, I will discuss within-subject ANOVA that focuses on repeated measures applied to the subjects of one group. The one-way ANOVA (Chapter 9) and the factorial ANOVA (Chapter 10) treatments of ANOVA focused on “between-subject” applications. That is, do sample groups differ on some outcome measure? As you recall, I explored the differences in health resulting from different sleep conditions (one-way ANOVA) and then extended this to include the differences between males and females (factorial ANOVA) by including the second factor (sex).

t-Test: Paired Two Sample for Means

	<i>Time1</i>	<i>Time2</i>
Mean	40.51	36.76
Variance	279.35	275.05
Observations	45.00	45.00
Pearson Correlation	0.94	
Hypothesized Mean Difference	0.00	
df	44.00	
t Stat	4.23	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.68	
P(T<=t) two-tail	0.00	
t Critical two-tail	2.02	

Figure 15.6 The T_{dep} findings from Excel paired two sample test.**TABLE 15.6** The Experimental Design

Pre Test Scores	Experimental Group Treatment	Post Test Scores
Pre Test Scores	Comparison Group(s)	Post Test Scores

EXPERIMENTAL DESIGNS

Table 15.6 shows the “classic” experimental design in which two groups (randomly selected in the “true” experiment and not in the “quasi-experiment”) are compared on posttest scores after some intervention. The experimental group receives an intervention, and the control or comparison group does not receive the intervention. *If the posttest scores differ between the groups, the implication is that the treatment is responsible.*

An example might be the factorial ANOVA (Chapter 10) in which I analyzed sleep conditions (factor 1) and sex (factor 2) as influences on health. To do this, I assumed that the experiment included randomization of subjects so that I did not have to include the pretests. The assumption is that if subjects are randomly chosen, the groups’ pretest measures will be equal and therefore will not need to be included in the analyses. I simply compared the posttest (learning) scores for men and women in the different noise conditions.

There is also value in knowing how the individual subjects of each group change across the time of the experiment. If the experimental treatment is effective, it should “cause” changes in individuals’ posttest scores compared to their pretest scores.

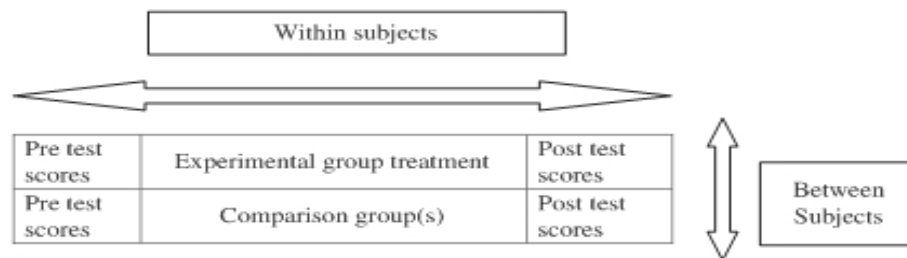


Figure 15.7 The mixed design that includes within-subject and between-group elements.

TABLE 15.7 Data for Within-Subject Study with Three Categories

	Time1	Time2	Time3
Subject 1	11	6	3
Subject 2	4	2	3
Subject 3	11	11	4
	Mean Time1	Mean Time2	Mean Time3

It is often difficult to achieve full randomization, so you cannot assume the pretest measures of different treatment groups (experimental and comparison groups) will be equal. Therefore, you need to take into account the differences that happen from pretest to posttest within both (all) treatment groups. *The experimental design that does not include full randomization therefore may include both within-subject and between-subject measures.* This is known as a “mixed design.” Figure 15.7 shows how this works.

In this chapter, I discuss the within-subject element of this design because it is often the case that research simply focuses on how one group of subjects change over time. To extend the example from the T_{dep} section earlier, this might mean conducting tissue test times *assessments three times for each subject* in the experiment. Table 15.7 shows how this design looks.

Since I now have three measures for each lab technician, I extend the T_{dep} test to a within-subject ANOVA. This procedure will detect differences for each lab technician among the three tissue test assessments. Since the three assessments are based on the same technicians, you must use a statistical procedure that factors out the sameness so that you can see what the resulting differences are among the three assessments. Will the training session result in declining times for the tissue test procedures across the three assessments?

POST FACTO DESIGNS

Within-subject procedures can be used with post facto designs as well since you are not necessarily assuming experimental manipulation. You might simply be interested

TABLE 15.8 The Within-Subject Example Data

Time1	Time2	Time3	Time1	Time2	Time3
11	6	3	41	35	40
4	2	3	62	47	40
11	11	4	29	29	40
18	9	6	55	50	40
30	23	7	46	49	41
7	12	8	39	38	41
12	10	9	41	40	41
42	42	9	45	46	44
32	20	12	49	58	47
17	5	18	41	35	48
24	24	20	54	56	49
46	37	21	50	53	49
26	28	22	52	58	51
28	31	27	50	49	52
44	39	30	56	57	52
44	30	32	62	60	56
38	38	33	57	39	57
45	39	34	54	48	62
35	34	36	79	70	62
45	36	36	64	62	62
39	31	37	47	41	63
40	37	38	64	50	75
48	40	40			

in whether a group of students' AP course assessments are consistent across a certain time period (e.g., semester) or if they are erratic for some reason. The key issue is how the same students (or matched group of students) compare to themselves across different data collection periods or conditions.

WITHIN-SUBJECT EXAMPLE

Table 15.8 shows the data I will use for this example. As you can see, it is the same data I used for the T_{dep} example with an additional assessment period for each lab technician (Time3). For the purposes of this example, I will consider the time periods as 1-month intervals with the training session taking place immediately after the Time1 assessment. Will the training approach produce consistent results (lower tissue test times) over the time periods?

USING SPSS® FOR WITHIN-SUBJECT DATA

For this example, I will not present the manual calculations but focus on the use of the statistical program to produce the findings. Excel does not have a straightforward method for calculating the one-way ANOVA_{ws}. I will therefore focus on the SPSS program which has detailed procedures for this design. I will discuss the procedure using the data in Table 15.8.

The output for ANOVA_{ws} is quite complex. Therefore, I present only the basic output to show how to interpret the primary findings.

Sphericity

One of the assumptions of ANOVA_{ws} is that the variances among the subjects' (in this case) three time periods be approximately equal. That is, the variance between Time1 and Time2 should be equal, the variance between Time2 and Time3 should be equal, and the variance between Time1 and Time3 should be equal. This is like the assumption I discussed for one-way ANOVA that was assessed by the Levene's test. With the ANOVA_{ws} you assess the variances to make sure they are equivalent by the "Mauchly's test of sphericity" and then proceed with the analyses depending on whether you have met or violated the assumptions.

THE SPSS® PROCEDURE

You create the specification for the ANOVA_{ws} through the "Analyze – General Linear Model – Repeated Measures" menu. This creates the window shown in Figure 15.8. As you can see, I specified at the top that the within-subject factor I will use is Time and that it has three measurements (specified in the "Number of Levels" window).

When I select "Define" the window shown in Figure 15.9 will appear. This summarizes my specification in the "Within-Subjects Variables" window. As you can see, there are several choices I can make to further specify the model on the right side of this window.

Choosing the "Contrasts" button allows the user to specify that the comparison of conditions is based on repeated measures. Figure 15.10 shows this "Repeated Measures: Contrasts" window. As you can see, I clicked on "Time" and then chose the

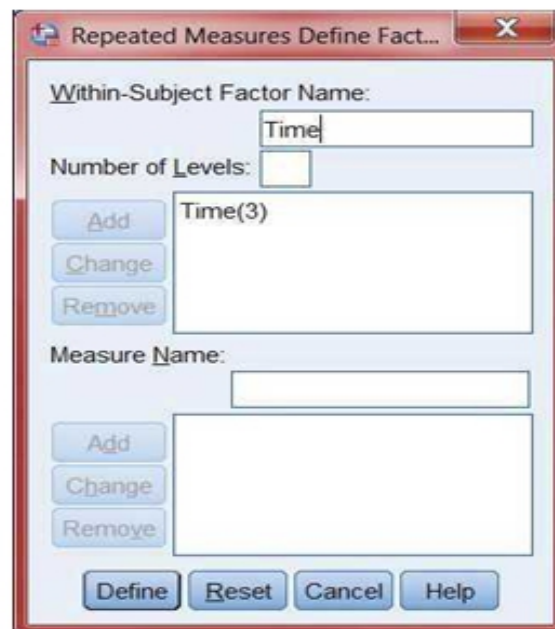


Figure 15.8 The SPSS® specification window for the ANOVA_{ws} procedure.

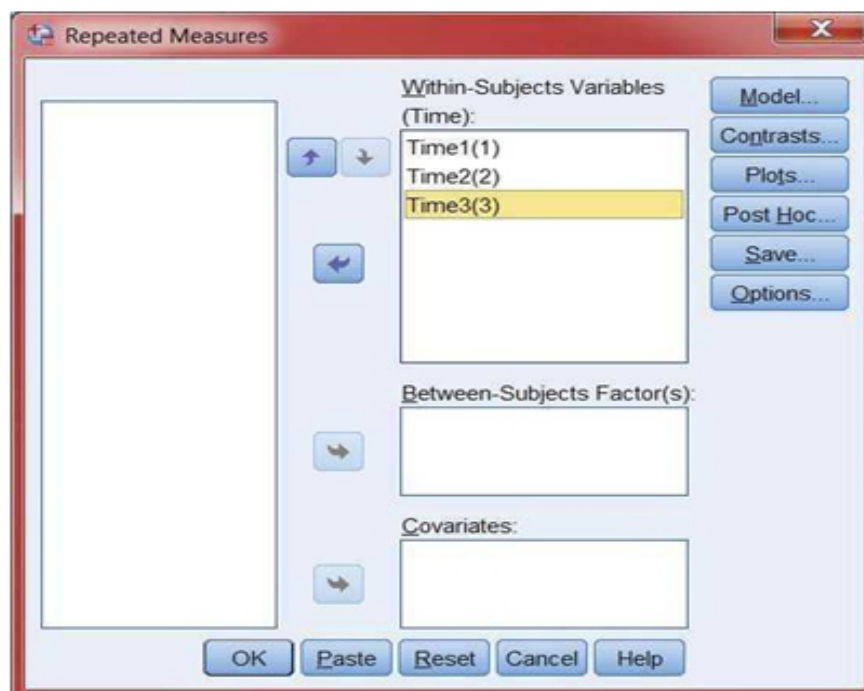


Figure 15.9 The SPSS® “Repeated Measures” window.

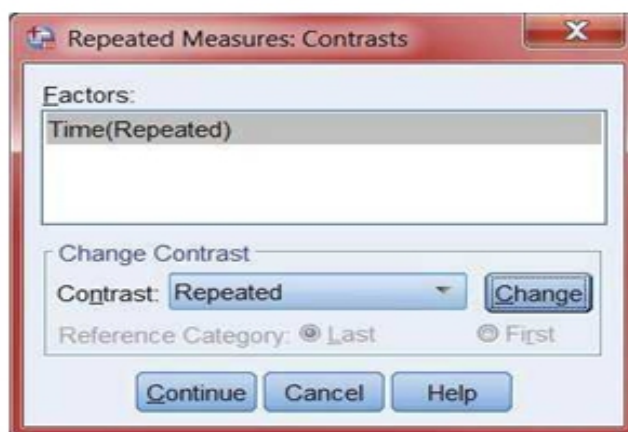


Figure 15.10 The “Contrasts” window for specifying repeated contrasts.

“Change Contrast” button which allowed me to choose “Repeated” and then clicking “Change” to make sure the change registered in the window. This is shown in Figure 15.10 where the “Factors” window shows “Time(Repeated).”

Returning to the main “Repeated Measures” window shown in Figure 15.9, you can choose the “Plots” button to specify Time on the “Horizontal Axis.” I will show this result in the following.

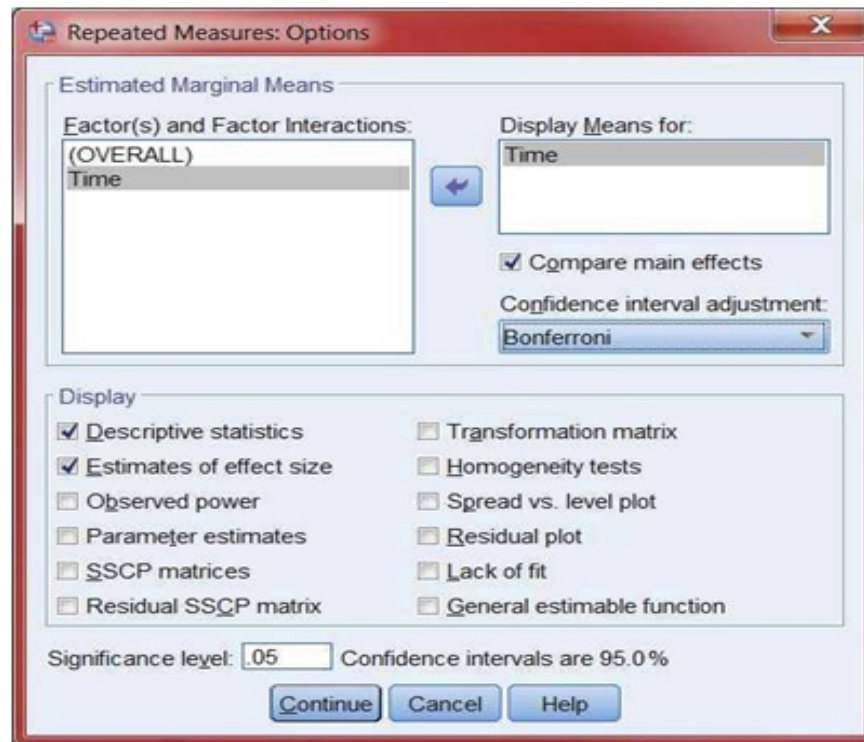


Figure 15.11 The “Options” menu for the $ANOVA_{ws}$ procedure.

When I make these selections, I next can choose the “Options” button in the main “Repeated Measures” window (shown in Figure 15.9). This creates the window shown in Figure 15.11. As you can see, I have chosen to “Display Means for” Time by moving it to the right side window. I also checked the “Compare main effects” box and specified the “Bonferroni” adjustment from the “Confidence interval adjustment” button just below the “Display Means” window. Under “Display” I checked the “Descriptive statistics” and “Estimates of effect size” boxes to show the appropriate results in the output.

THE SPSS® OUTPUT

In this book, I will provide only a basic look at the SPSS $ANOVA_{ws}$ analyses and how to create general interpretations. As the research design increases in complexity, however, so does the design specification and output. You will need to consult additional sources if you choose to elaborate on the one-way $ANOVA_{ws}$ design shown here.

Figure 15.12 shows the descriptive table that includes the means of the three time conditions for the 45 subjects. You can see that the SD measures are generally equal, but the SD measure for Time3 is slightly larger. When you test the equivalence in variances (through the Mauchly’s test of sphericity), you will assess whether this difference will be problematic.

Descriptive Statistics			
	Mean	Std. Deviation	N
Time1	40.51	16.714	45
Time2	36.76	16.585	45
Time3	35.49	18.817	45

Figure 15.12 The “Descriptive Statistics” output.

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
Time	.694	15.697	2	.000	.766	.788	.500

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

a. Design: Intercept

Within Subjects Design: Time

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.

Figure 15.13 The Mauchly's test of sphericity results.

The output for Mauchly's test of sphericity is shown in Figure 15.13. The critical value to look for is the “Sig.” value shown in the middle of the table. In this case, the results show that there are significant differences among the variance measures of the three time conditions. Thus, there is a violation of the sphericity condition. These can be very sensitive tests, so I will take note of the violation in the analysis of the remaining output.

The Omnibus Test

Figure 15.14 shows the output for the within-subject effects of time. The first “group” of findings under time shows four different rows of output. The first “Sphericity Assumed” is like the one-way ANOVA F test in which there are equal variances. As you can see, this test is significant ($F = 7.830$, $p < 0.01$). If there are no sphericity violations, you would use this value to make the conclusions.

However, since there is a violation of sphericity, you would need to use an “adjusted” F test value. As you can see from Figure 15.14, SPSS provides three such adjusted tests. The “Greenhouse–Geisser” test is shown in the second row and is a commonly used conservative test. As you can see, this F value is the same, but the significance level is slightly different due to the adjustment for sphericity. The result is the same however. Time is a significant repeated measure condition ($F = 7.830$, $p = 0.002$).

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	613.970	2	306.985	7.830	.001	.151
	Greenhouse-Geisser	613.970	1.532	400.870	7.830	.002	.151
	Huynh-Feldt	613.970	1.576	389.624	7.830	.002	.151
	Lower-bound	613.970	1.000	613.970	7.830	.008	.151
Error(Time)	Sphericity Assumed	3450.030	88	39.205			
	Greenhouse-Geisser	3450.030	67.390	51.195			
	Huynh-Feldt	3450.030	69.335	49.759			
	Lower-bound	3450.030	44.000	78.410			

Figure 15.14 The within-subject effects output for Time.

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.359	12.020 ^a	2.000	43.000	.000	.359
Wilks' lambda	.641	12.020 ^a	2.000	43.000	.000	.359
Hotelling's trace	.559	12.020 ^a	2.000	43.000	.000	.359
Roy's largest root	.559	12.020 ^a	2.000	43.000	.000	.359

Each F tests the multivariate effect of Time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Figure 15.15 The effect size output for $ANOVA_{WS}$.

Effect Size

Figure 15.15 shows the multivariate test results. These reveal the *effect size* findings for this study and are not impacted by sphericity. I have not discussed these so far, but many researchers use “Wilks’ lambda” as the effect size measure for $ANOVA_{WS}$ in this situation. Wilks’ lambda (Λ) is based on measuring unexplained variance, so the smaller the value, the stronger the effect.

As you can see from Figure 15.15, $\Lambda = 0.641$ and is significant ($p < 0.0001$). You will also notice that in this test, Λ is complimentary to partial eta squared which is not based on error but on the size of the regression (i.e., partial $\eta^2 = 1 - \Lambda$). As you can see, partial $\eta^2 = 0.359$ and Wilks’ lambda = 0.641, suggesting that the impact of the training on tissue test results across time conditions is considered large (using the criteria I suggested in Chapter 10).

Post Hoc Analyses

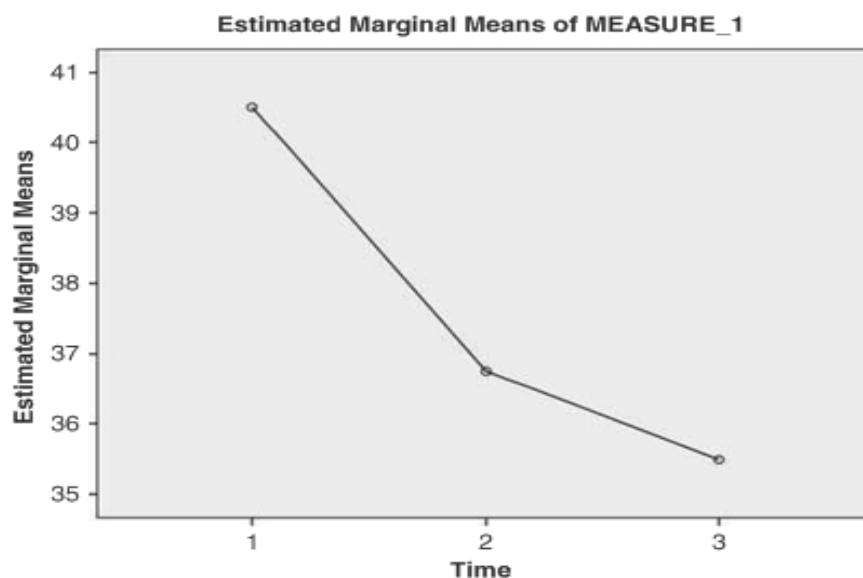
Just as I showed with the ANOVA tests in Chapter 9, you must perform post hoc analyses when the omnibus test is significant. Figure 15.16 shows the pairwise comparisons output that indicates which time conditions differ. As you can see, all the comparisons are significantly different except the difference between Time2 and Time3.

Pairwise Comparisons

Measure: MEASURE_1

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	3.756 ^a	.887	.000	1.968	5.543
	3	5.022 ^a	1.453	.001	2.094	7.951
2	1	-3.756 ^a	.887	.000	-5.543	-1.968
	3	1.267	1.526	.411	-1.809	4.342
3	1	-5.022 ^a	1.453	.001	-7.951	-2.094
	2	-1.267	1.526	.411	-4.342	1.809

Based on estimated marginal means

^a. The mean difference is significant at the .05 level.^b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).**Figure 15.16** The post hoc output for the study.**Figure 15.17** The comparison plot for the time conditions.

You can see that the differences among the means of the time conditions are large except for Time2–Time3 (1.267). The plot in Figure 15.17 provides visual comparisons in the plotting of the means. As you can see, there are large differences (less time for tissue tests) between Time1 and Time2, and between Time1 and Time3, but the difference between Time2 and Time3 is smaller and not significant.

The Interpretation

Taking all the output into account, it appears that training lab technicians does lower their time for tissue tests across the time conditions in this hypothetical study. Each

time period indicates less time for the tests, especially after the training. The primary differences are between the initial assessment (Time1) and both of the subsequent assessments (Time2 and Time3). The effects are not different between the latter two assessments (Time2–Time3).

NONPARAMETRIC STATISTICS

In past chapters, I presented the nonparametric counterparts of some parametric tests. Thus, for the independent sample T test, I discussed the Mann–Whitney U test, and I presented a section on the Kruskal–Wallis test when I discussed the one-way ANOVA procedure.

Both of the repeated measures tests in the current chapter have nonparametric counterparts as well. With *two* related (dependent) samples of ordinal data, the researcher would use the Wilcoxon test. If there are *more than two* dependent samples of less than interval data, the researcher would use the Friedman test. Both of these are available in SPSS and can be accessed with the same “analyze” menu I demonstrated with the Mann–Whitney U and Kruskal–Wallis tests.

TERMS AND CONCEPTS

Dependent samples:	Groups in which the members are structurally related, such as using the same group of subjects twice in an experiment (pre–post) or using “matched groups.” Also known as paired samples, repeated measures, and within-subject measures.
Friedman test:	A nonparametric test used with more than two dependent samples (repeated measures) with ordinal data.
Greenhouse–Geisser test:	This is one of the tests used by SPSS to “adjust” values affected by sphericity.
Sphericity:	The assumption with repeated measures tests (like ANOVA_{WS}) that the variance of group differences is equal.
Wilcoxon test:	A nonparametric test used with two dependent samples (repeated measures) with ordinal data.
Wilks’ lambda:	An effect size measure based on the amount of unexplained variance. Small results are considered stronger than large results.