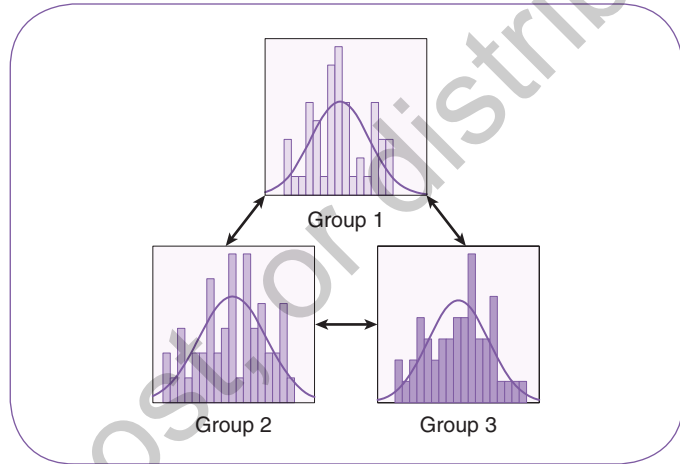


CHAPTER 6

ANOVA and Kruskal-Wallis Test

To compare more than two groups of continuous variables, run an **ANOVA** or **Kruskal-Wallis test**.



Three is a magic number.

—Bob Dorough

LEARNING OBJECTIVES

Upon completing this chapter, you will be able to:

- Determine when it is appropriate to run an ANOVA test
- Verify that the data meet the criteria for ANOVA processing: normality, n , and homogeneity of variance
- Order an ANOVA test with graphics
- Select an appropriate ANOVA post hoc test: Tukey or Sidak
- Derive results from the *Descriptives* and *Multiple Comparisons* tables
- Calculate the unique pairs formula
- Resolve the hypotheses
- Know when and how to run and interpret the Kruskal-Wallis test
- Document the results in plain English



VIDEOS

The videos for this chapter are **Ch 06 - ANOVA.mp4** and **Ch 06 - Kruskal-Wallis Test.mp4**. These videos provide overviews of these tests, instructions for carrying out the pretest checklist, running the tests, and interpreting the results of each test using the data set **Ch 06 - Example 01 - ANOVA and Kruskal-Wallis.sav**.



LAYERED LEARNING

The t test and ANOVA (analysis of variance) are so similar that this chapter will use the same example and the same 10 exercises used in Chapter 5 (“ t Test and Mann-Whitney U Test”); the only difference is that the data sets have been enhanced to include a third or fourth group. If you are proficient with the t test, you are already more than halfway there to comprehending ANOVA. The only real differences between the t test and ANOVA are in ordering the test run and interpreting the test results; several other minor differences will be pointed out along the way.

That being said, let us go into the expanded example, drawn from Chapter 5, which involved timing how long it takes to put together the Able Chair using different types of assembly instructions: Group 1 (Text only), Group 2 (Text with illustrations), and now a third group, Group 3 (Video). The ANOVA test will reveal which (if any) of these three types of assembly instructions statistically significantly outperforms the others in terms of how long it takes to assemble the Able Chair.



OVERVIEW—ANOVA

The ANOVA test is similar to the t test, except that whereas the t test compares two groups of continuous variables with each other, the ANOVA test can compare three or more groups with one another.

In cases where the three pretest criteria are not satisfied for the ANOVA, the Kruskal-Wallis test, which is conceptually similar to the ANOVA, is the better option; this alternative test is explained near the end of this chapter.

Example

The manufacturers of the ergonomic Able Chair want to determine the most efficient assembly instructions for their consumers: (1) Text only, (2) Text with illustrations, or (3) Video.

Research Question

Which instruction method produces the quickest consumer assembly of the Able Chair?

Groups

A researcher recruits a total of 105 participants to assemble the Able Chair. Participants will be scheduled to come to the research center one at a time. Upon arriving, each participant will be assigned to one of three groups (first person, Group 1; second person, Group 2; third person, Group 3; fourth person, Group 1, and so on). Those assigned to Group 1 will be issued text-only assembly instructions, those assigned to Group 2 will be issued assembly instructions containing text with illustrations, and those assigned to Group 3 will be issued a video demonstrating how to assemble the chair.

Procedure

Each participant will be guided to a room where there is a new Able Chair in its regular package, a screwdriver (the only tool required), and the assembly instructions. The researcher will use a stopwatch to time how long it takes each participant to assemble the chair.

Hypotheses

The null hypothesis (H_0) is phrased to anticipate that the treatment (inclusion of the illustrations or use of video) fails to shorten the assembly time, indicating that on average, participants who are issued these alternate forms of assembly instructions will take just as long to assemble the chair as those who are issued the text-only instructions; in other words, there is no difference between the assembly times among these groups. The alternative hypothesis (H_1) states that on the average, at least one group will outperform another group:

H_0 : There is no difference in assembly time across the groups.

H_1 : At least one group outperforms another in terms of assembly time.

Admittedly, H_1 is phrased fairly broadly. The post hoc *Multiple Comparisons* table, which is covered in the **Results** section, will identify which teaching method(s), if any, outperform which.



Data Set

Use the following data set: **Ch 06 - Example 01 - ANOVA and Kruskal-Wallis Test.sav**.

Notice that this data set has 105 records; the first 70 records (rows) are the same as the t test and Mann-Whitney U test example data set used in Chapter 5 (records 71 through 105 are new).

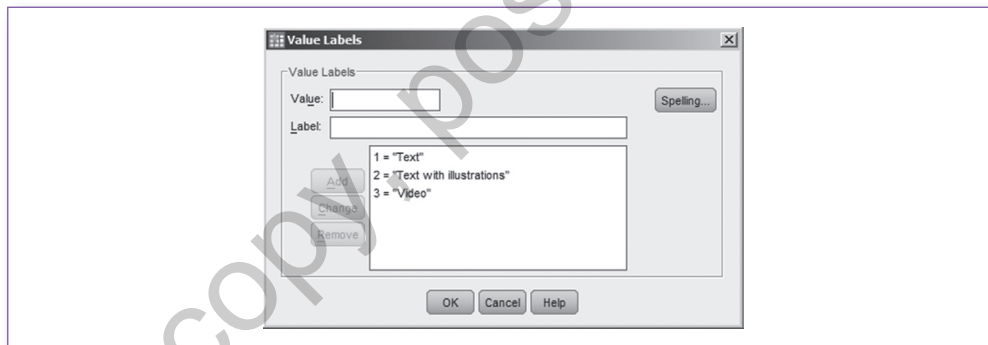
Codebook

Variable:	group
Definition:	Group assignment
Type:	Categorical (1 = Text, 2 = Text with illustrations, 3 = Video)
Variable:	time
Definition:	Minutes required to assemble the Able Chair
Type:	Continuous

NOTE: In this data set, records (rows) 1 through 35 are for Group 1 (Text), records 36 through 70 are for Group 2 (Text with illustrations), and records 71 through 105 are for Group 3 (Video). The data are arranged this way just for visual clarity; the order of the records has no bearing on the statistical results.

If you go to the *Variable View* and open the *Value Labels* window for the variable *group*, you will see that the label “Video” for the third group has been assigned to the value 3 (Figure 6.1).

Figure 6.1 Value labels for a three-group ANOVA.

**Pretest Checklist****ANOVA Pretest Checklist**

- 1. Normality^a
- 2. n quota^b
- 3. Homogeneity of variance^b

a. Run prior to ANOVA test.

b. Results produced upon ANOVA test run.

The statistical pretest checklist for the ANOVA is similar to the t test—(a) **normality**, (b) n , and (c) **homogeneity of variance**—except that you will assess the data for more than two groups.



Pretest Checklist Criterion 1—Normality

Check for normality by inspecting the histogram with a normal curve for each of the three groups. Begin by using the *Select Cases* icon to select the records pertaining to the Text group; the selection criterion would be *group* = 1. Next, run a histogram (with normal curve) on the variable *time*. For more details on this procedure, refer to the section “SPSS—Descriptive Statistics: Continuous Variable (*Age*) Select by Categorical Variable (*Gender*)—Female or Male Only” in Chapter 4; see the star (★) icon on page 81.

Then repeat the process for the Text with illustrations group (*group* = 2), and finally, repeat the process a third time for the Video group (*group* = 3).

This will produce three histograms with normal curves—one for the scores in the Text group, a second for the scores in the Text with illustrations group, and a third for the Video group. The histograms should resemble the graphs shown in Figures 6.2, 6.3, and 6.4.

Figure 6.2 Histogram of score for Group 1: Text.

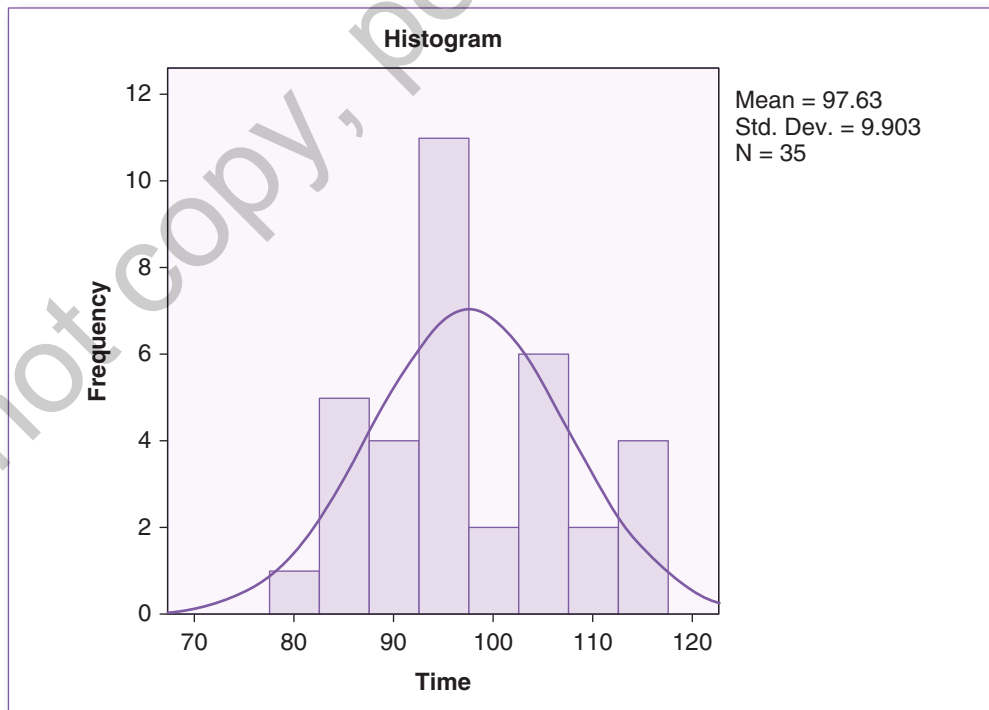
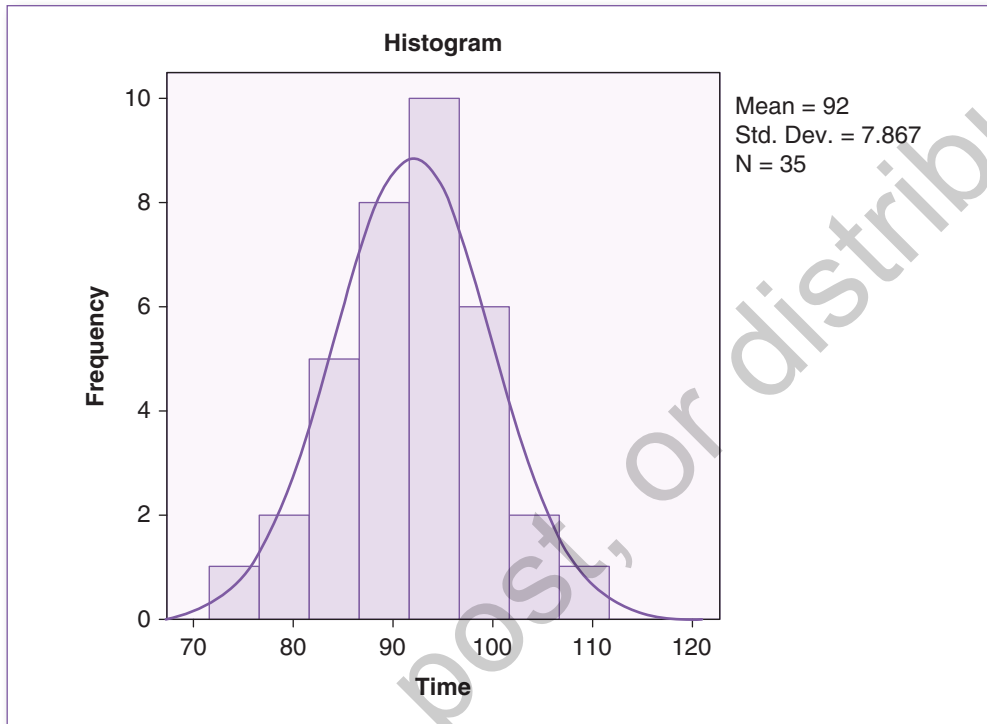


Figure 6.3 Histogram of score for Group 2: Text with illustrations.

As we read these three histograms, our focus is primarily on the *normality of the curve*, as opposed to the characteristics of the individual bars. Although the height and width of each curve are unique, we see that each is bell shaped and shows good symmetry, with no substantial *skewing*. On the basis of the inspection of these three figures, we would conclude that the criteria of *normality* are satisfied for all three groups.

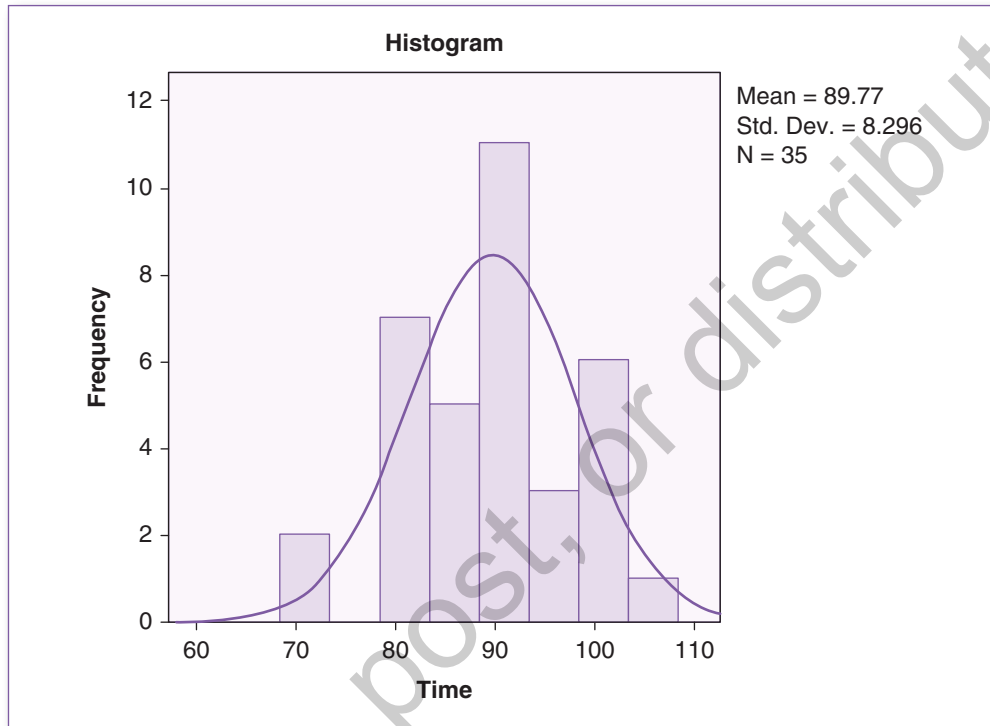
Next, (re)activate all records for further analysis; you can either delete the temporary variable *filter_\$* or click on the *Select Cases* icon and select the *All cases* button. For more details on this procedure, please refer to the section “SPSS—(Re)Selecting All Variables” in Chapter 4; see the star (★) icon on page 85.



Pretest Checklist Criterion 2—*n* Quota

Again, as with the *t* test, technically, you can run an ANOVA test with an *n* of any size in each group, but when the *n* is at least 30 in each group, the ANOVA is considered more robust. The *ns* will be part of the output produced by the **Test Run** procedure; we will revisit this criterion in the **Results** section.

Figure 6.4 Histogram of score for Group 3: Video.



Pretest Checklist Criterion 3—Homogeneity of Variance

Homogeneity pertains to sameness; the homogeneity of variance criterion involves checking that the variances of the two groups are similar to each other. As a rule of thumb, homogeneity of variance is likely to be achieved if the variance (standard deviation squared) from one group is not more than twice the variance of the other group. In this case, the variance for *time* in the Text group is 82.63 (derived from Figure 6.2: $9.0903^2 = 82.6336$), the variance for *time* in the Text with illustrations group is 61.89 (derived from Figure 6.3: $7.867^2 = 61.8897$), and the variance for *time* in the Video group is 61.89 (derived from Figure 6.4: $8.296^2 = 68.8236$). When looking at the *time* variances from these three groups (82.63, 61.89, and 68.82), clearly none of these figures is more than twice any of the others, so we would expect that the homogeneity of variance test would pass.

In SPSS, the homogeneity of variance test is an option selected during the actual run of the ANOVA test. If the homogeneity of variance test renders a significance (p) value that is greater than .05, this suggests that there is no statistically significant difference between the variance of one group and that of the other group. This would mean that

the data pass the homogeneity of variance test. The notion of the p value will be discussed in detail in the **Results** section in this chapter, when we examine the findings produced by the ANOVA test.

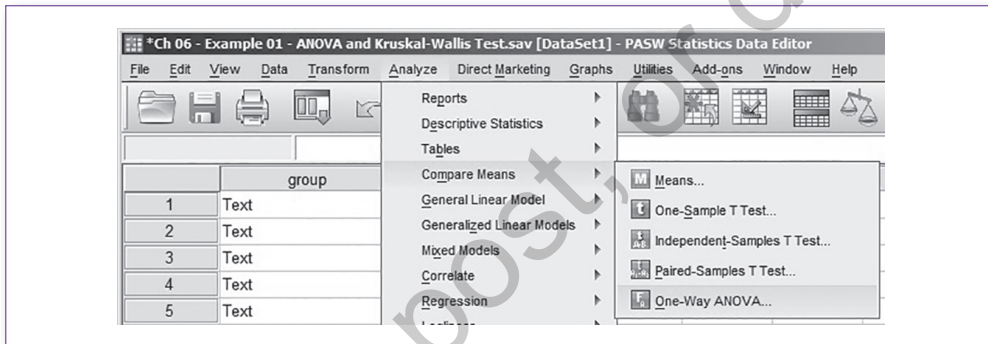


Test Run

To run an ANOVA test (for the most part, this is t test déjà vu time):

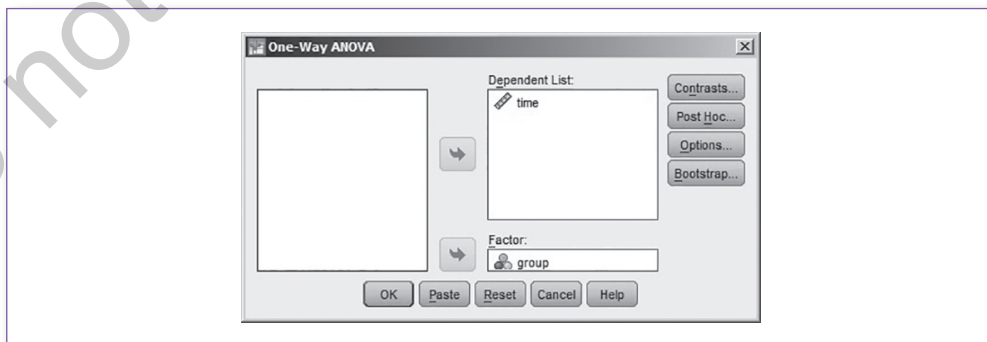
1. On the main screen, click on *Analyze, Compare Means, One-Way ANOVA* (Figure 6.5).

Figure 6.5 Running an ANOVA test.



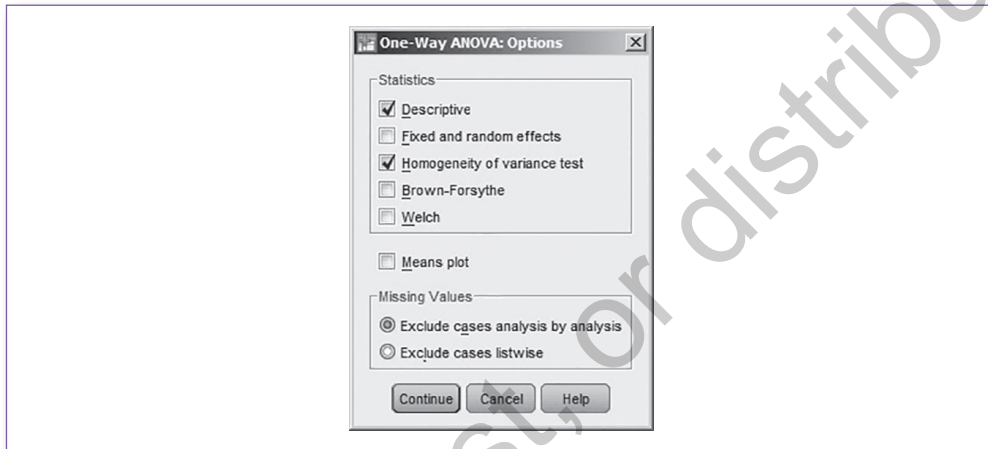
2. In the *One-Way ANOVA* window, move the continuous variable that you wish to analyze (*score*) into the *Dependent List* panel, and move the variable that contains the categorical variable that specifies the group (*group*) into the *Factor* panel (Figure 6.6).

Figure 6.6 The *One-Way ANOVA* window.



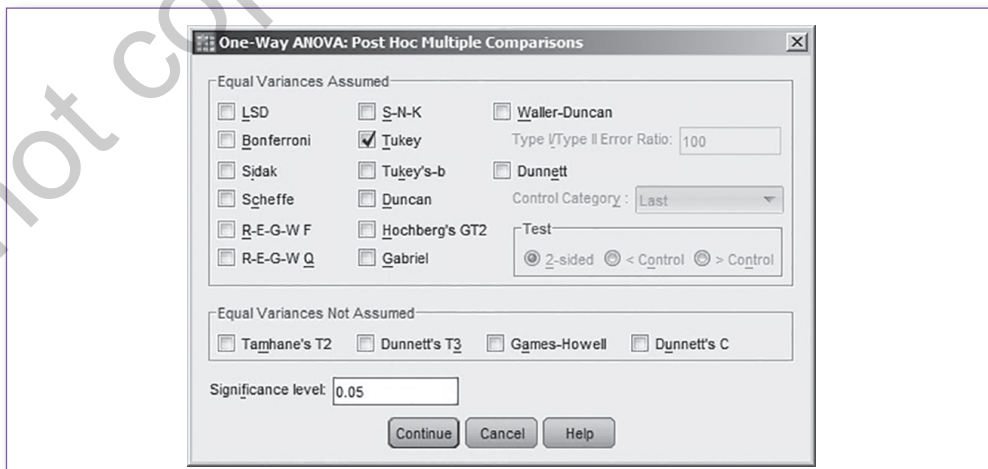
3. Click on the *Options* button. In the *One-Way ANOVA: Options* window, check *Descriptive* and *Homogeneity of variance test*, then click on the *Continue* button (Figure 6.7). This will take you back to the *One-Way ANOVA* window.

Figure 6.7 The *One-Way ANOVA: Options* window.



4. Click on the *Post Hoc* button.
5. This will take you to the *One-Way ANOVA: Post Hoc Multiple Comparisons* window (Figure 6.8).

Figure 6.8 The *One-Way ANOVA: Post Hoc Multiple Comparisons* window.



6. If you were to run the ANOVA test without selecting a post hoc test, all it would return is a single p value; if that p value indicates statistical significance, that would tell you that somewhere among the groups processed, the mean for at least one group is statistically significantly different from the mean of at least one other group, but it would not tell you specifically which group is different from which. The post hoc test produces a table comparing the mean of each group with the mean of every other group, along with the p value for each pair of comparisons. This will become clearer in the **Results** section when we read the post hoc *Multiple Comparisons* table.

As for which post hoc test to select, there are a lot of choices. We will focus on only two options: *Tukey* and *Sidak*. *Tukey* is appropriate when each group has the same n ; in this case, each group has an n of 35, so check the *Tukey* checkbox, then click on the *Continue* button (this will take you back to the *One-Way ANOVA* window [Figure 6.6]). If the groups had *different ns* (e.g., $n[\text{Group 1}] = 40$, $n[\text{Group 2}] = 55$, $n[\text{Group 3}] = 36$), then the *Sidak* post hoc test would be appropriate. If you do not know the n of each group in advance, just select either *Tukey* or *Sidak* and observe the ns on the resulting report; if you chose incorrectly, go back and rerun the analysis using the appropriate post hoc test.

ANOVA Post Hoc Summary

- If all groups have the same n , select *Tukey*.
- If the groups have different ns , select *Sidak*.

7. In the *One-Way ANOVA* window (Figure 6.6), click on the *OK* button, and the ANOVA test will process.

Results

Pretest Checklist Criterion 2— n Quota

Table 6.1 shows that each group has an n of 35. This satisfies the n assumption, indicating that the ANOVA test becomes more robust when the n for each group is at least 30.

Pretest Checklist Criterion 3—Homogeneity of Variance

As for the final item on the pretest checklist, Table 6.2 shows that the homogeneity of variance test produced a Sig. (p) value of .122; since this is greater than the α level of .05, this tells us that there are *no statistically significant differences among the variances of the score variable for the three groups analyzed*. In other words, the variances for *score* are

Table 6.1 Descriptive Statistics (n) of *time* for the Text, Text With Illustrations, and Video Groups.

Descriptives								
time	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					Text	35		
Text with illustrations	35	92.00	7.867	1.330	89.30	94.70	74	110
Video	35	89.77	8.296	1.402	86.92	92.62	71	105
Total	105	93.13	9.264	.904	91.34	94.93	71	115

similar enough among the three groups (Text, Text with illustrations, and Video), hence we would conclude that the criterion of homogeneity of variance has been satisfied.

Table 6.2 Homogeneity of Variance Test Results.

Test of Homogeneity of Variances				
time	Levene Statistic	df1	df2	Sig.
	2.147	2	102	.122

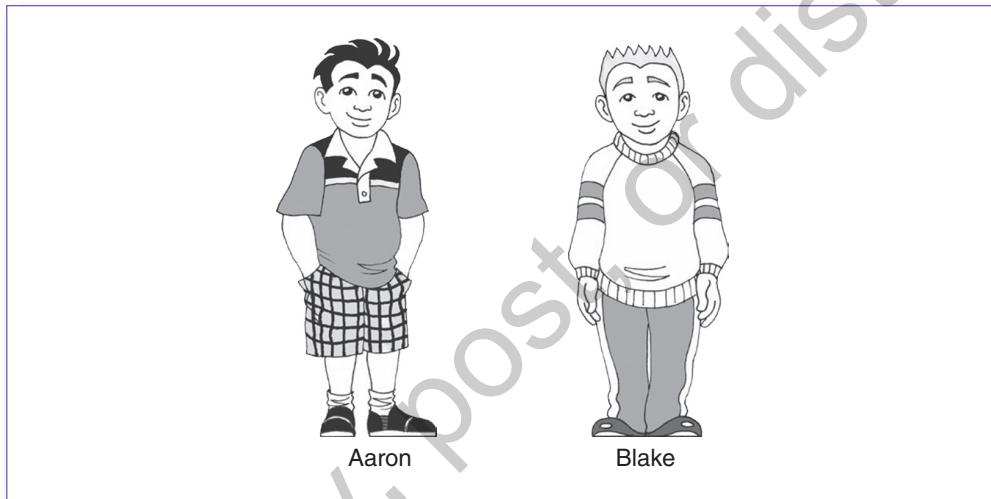
Next, we look at the ANOVA table (Table 6.3) and find a Sig. (p) value of .001; since this is less than the α level of .05, this tells us that there is a statistically significant difference among the (three) group means for *score*, but unlike reading the results of the t test, we are not done yet.

Table 6.3 ANOVA Test Results Comparing *time* of the Text, Text With Illustrations, and Video Groups.

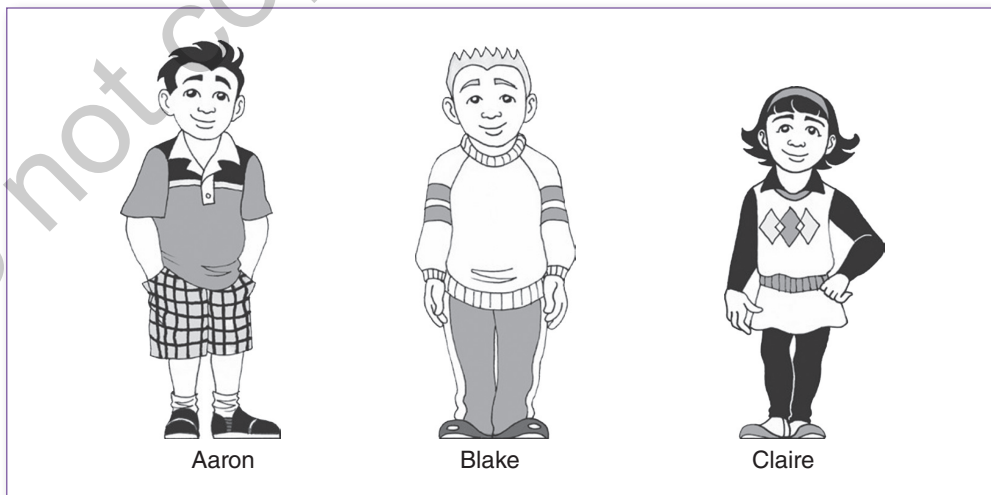
ANOVA					
time	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1147.790	2	573.895	7.526	.001
Within Groups	7778.343	102	76.258		
Total	8926.133	104			

Remember that in the realm of the t test, there are only *two* groups involved, so interpreting the p value is fairly straightforward: If $p \leq .05$, there is no question as to which group is different from which—clearly, the mean of Group 1 is statistically significantly different from the mean of Group 2. However, when there are *three or more groups*, we need more information to determine which group is different from which; that is what the post hoc test answers.

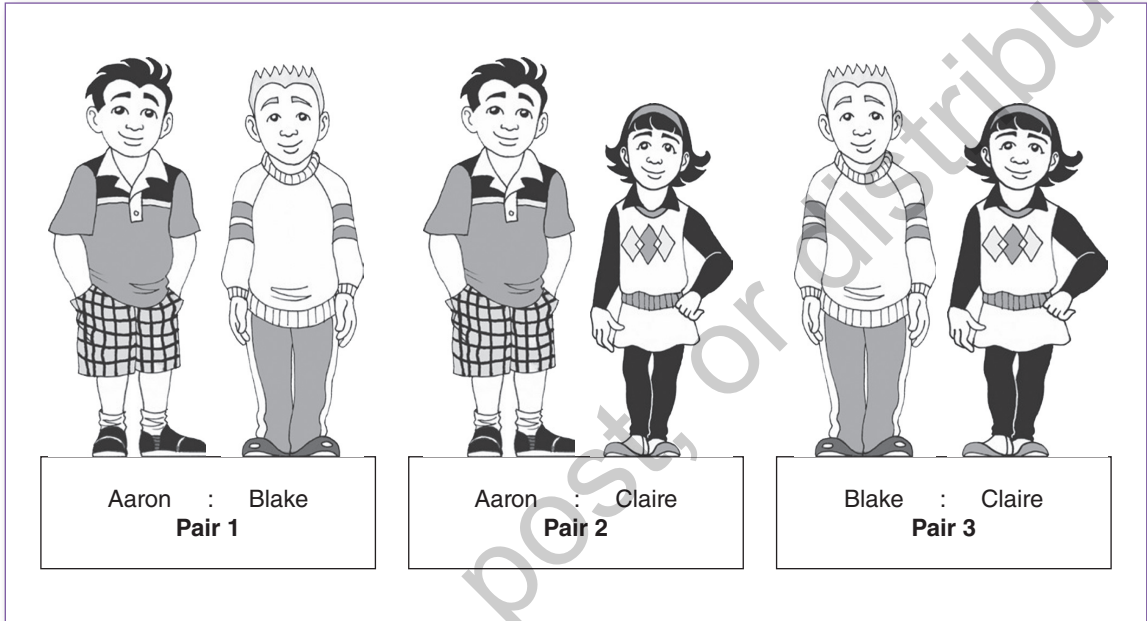
Consider this: Suppose you have two kids, Aaron and Blake; you are in the living room, and someone calls out from the den, “The kids are fighting again!” Since there are only two kids, you immediately know that the fight is between Aaron and Blake; this is akin to the t test, which involves comparing the means of two groups.



Now suppose you have three kids—Aaron, Blake, and Claire:



This time when the voice calls out, “The kids are fighting again!” you can no longer simply know that the fight is between Aaron and Blake; when there are three kids, you need more information. Instead of just one possibility, there are now three possible pairs of fighters:



Back to our example: The ANOVA table (Table 6.3) produced a statistically significant p value (Sig. = .001), which indicates that there is a statistically significant difference detected somewhere among the three groups (“The kids are fighting!”); the post hoc table will tell us precisely which pairs are statistically significantly different from each other (which pair of kids is fighting). Specifically, it will reveal which group(s) outperformed which.

This brings us to the (Tukey post hoc) *Multiple Comparisons* table (Table 6.4). As with the three kids fighting, in this three-group design, there are three possible pairs of comparisons we can assess in terms of (mean) *score* for the groups:

Group 1 : Group 2 97.63 : 92.00 Pair 1	Group 1 : Group 3 97.63 : 89.77 Pair 2	Group 2 : Group 3 92.00 : 89.77 Pair 3
---	---	---

Means are shown for Group 1 (Text), Group 2 (Text with illustrations), and Group 3 (Video).

We will use Table 6.1 (*Descriptives*) and Table 6.4 (*Multiple Comparisons*) to analyze the ANOVA test results. Table 6.1 lists the mean time for each of the three groups:

$\mu(\text{Text}) = 97.63$, $\mu(\text{Text with illustrations}) = 92.00$, and $\mu(\text{Video}) = 89.77$. We will assess each of the three pairwise score comparisons separately.

Comparison 1—Text : Text With Illustrations

Table 6.4 first compares the mean time for the Text group with the mean score for the Text with illustrations group, which produces a Sig. (p) of .022. Since p is less than the .05 α level, this tells us that for *time*, there is a statistically significant difference between Text ($\mu = 97.63$) and Text with illustrations ($\mu = 92.00$).

Table 6.4 ANOVA Post Hoc *Multiple Comparisons* Table Shows a Statistically Significant Difference Between the Text and Text With Illustrations Groups ($p = .022$).

Multiple Comparisons						
time						
Tukey HSD						
(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Text	Text with illustrations	5.629 ^a	2.087	.022	.66	10.59
	Video	7.857 ^a	2.087	.001	2.89	12.82
Text with illustrations	Text	-5.629 ^a	2.087	.022	-10.59	-.66
	Video	2.229	2.087	.536	-2.74	7.19
Video	Text	-7.857 ^a	2.087	.001	-12.82	-2.89
	Text with illustrations	-2.229	2.087	.536	-7.19	2.74

*. The mean difference is significant at the 0.05 level.

Comparison 2—Text : Video

The second comparison in Table 6.5 is between Text and Video, which produces a Sig. (p) of .001. Since p is less than the .05 α level, this tells us that for *time*, there is a statistically significant difference between Text ($\mu = 97.63$) and Video ($\mu = 89.77$).

Table 6.5 ANOVA Post Hoc *Multiple Comparisons* Table Shows a Statistically Significant Difference Between the Text and Video Groups ($p = .001$).

Multiple Comparisons						
time						
Tukey HSD						
(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Text	Text with illustrations	5.629 ^a	2.087	.022	.66	10.59
	Video	7.857 ^a	2.087	.001	2.89	12.82
Text with illustrations	Text	-5.629 ^a	2.087	.022	-10.59	-.66
	Video	2.229	2.087	.536	-2.74	7.19
Video	Text	-7.857 ^a	2.087	.001	-12.82	-2.89
	Text with illustrations	-2.229	2.087	.536	-7.19	2.74

*. The mean difference is significant at the 0.05 level.

Comparison 3—Text With Illustrations : Video

The third comparison in Table 6.6 is between Text with illustrations and Video, which produces a Sig. (*p*) of .536. Since *p* is greater than the .05 α level, this tells us that for *time*, there is no statistically significant difference between Text with illustrations ($\mu = 92.00$) and Video ($\mu = 89.77$).

Table 6.6 ANOVA Post Hoc *Multiple Comparisons* Table Shows No Statistically Significant Difference Between Text With Illustrations and Video (*p* = .536).

Multiple Comparisons						
(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Text	Text with illustrations	5.629*	2.087	.022	.66	10.59
	Video	7.857*	2.087	.001	2.89	12.82
Text with illustrations	Text	-5.629*	2.087	.022	-10.59	-.66
	Video	2.229	2.087	.536	-2.74	7.19
Video	Text	-7.857*	2.087	.001	-12.82	-2.89
	Text with illustrations	-2.229	2.087	.536	-7.19	2.74

*. The mean difference is significant at the 0.05 level.

This concludes the analysis of the *Multiple Comparisons* (post hoc) table. You have probably noticed that we skipped analyzing half of the rows; this is because there is a double redundancy among the figures in the Sig. column. This is the kind of double redundancy you would expect to see in a typical two-dimensional table. For example, in a multiplication table, you would see two 32's in the table, because $4 \times 8 = 32$ and $8 \times 4 = 32$. Similarly, the Sig. column of the *Multiple Comparisons* table contains two *p* values of .022: one comparing Text to Text with illustrations and the other comparing Text with illustrations to Text (Table 6.7). In addition, there are two *p* values of .001 (Text : Video and Video : Text) and two *p* values of .536 (Text with illustrations : Video and Video : Text with illustrations).

The ANOVA test can process any number of groups, provided the pretest criteria are met. As the number of groups increases, the number of (multiple) pairs of comparisons increases as well (see Table 6.8).

You can easily calculate the number of (unique) pairwise comparisons the post hoc test will produce:

Table 6.7

ANOVA Post Hoc *Multiple Comparisons* Table Containing Double-Redundant Sig. (*p*) Values—Text : Text with illustrations Produces the Same *p* Value as Text with illustrations : Text (*p* = .022).

Multiple Comparisons						
time						
Tukey HSD						
(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Text	Text with illustrations	5.629 ^a	2.087	.022	.66	10.59
	Video	7.857 ^a	2.087	.001	2.89	12.82
Text with illustrations	Text	-5.629 ^a	2.087	.022	-10.59	-.66
	Video	2.229	2.087	.536	-2.74	7.19
Video	Text	-7.857 ^a	2.087	.001	-12.82	-2.89
	Text with illustrations	-2.229	2.087	.536	-7.19	2.74

*. The mean difference is significant at the 0.05 level.

Table 6.8

Increasing the Number of Groups Substantially Increases ANOVA Post Hoc Multiple Comparisons.

<p>2 Groups Renders 1 Comparison</p> <p>$G_1:G_2$</p>	<p>3 Groups Renders 3 Comparisons</p> <p>$G_1:G_2$ $G_2:G_3$ $G_1:G_3$</p>	<p>4 Groups Renders 6 Comparisons</p> <p>$G_1:G_2$ $G_2:G_3$ $G_3:G_4$ $G_1:G_3$ $G_2:G_4$ $G_1:G_4$</p>
--	---	--

NOTE: G = group.



Unique Pairs Formula

G = Number of groups

$$\text{Number of ANOVA post hoc unique pairs} = G! \div [2 \times (G - 2)!]$$

The above formula uses the *factorial* function, denoted by the exclamation mark (!). If your calculator does not have a factorial (!) button, you can calculate it manually: Simply multiply all of the integers between 1 and the specified number. For example: $3! = 1 \times 2 \times 3$, which equals 6.

H₀

Hypothesis Resolution

To clarify the hypothesis resolution process, it is helpful to organize the findings in a table and use asterisks to flag statistically significant difference(s) (Table 6.9).

NOTE: SPSS does not generate this table (Table 6.9) directly; you can assemble this table by gathering the means from the *Descriptives* table (Table 6.1) and the p values from the Sig. column in the *Multiple Comparisons* table (Table 6.4).

Table 6.9 Results of ANOVA for *time*.

Groups	p
$\mu(\text{Text}) = 97.63 : \mu(\text{Text with illustrations}) = 92.00$.022*
$\mu(\text{Text}) = 97.63 : \mu(\text{Video}) = 89.77$.001*
$\mu(\text{Text with illustrations}) = 92.00 : \mu(\text{Video}) = 89.77$.536

*Statistically significant difference detected between groups ($p \leq .05$).

With this results table assembled, we can now revisit and resolve our pending hypotheses, which focus on determining the best type of instructions for assembling the Able Chair. To finalize this process, we will assess each hypothesis per the statistics contained in Table 6.9.

REJECT H₀: There is no difference in assembly time across the groups.

ACCEPT H₁: At least one group outperforms another in terms of assembly time.

Since we discovered a statistically significant difference among at least one pair of the comparisons of assembly instructions, we reject H₀ and accept H₁. Specifically, when it comes to the quickest assembly time, Text with illustrations outperformed Text ($p = .022$), and Video also outperformed Text ($p = .001$).

Incidentally, if all of the pairwise comparisons had produced p values that were greater than .05, we would have accepted H₀ and rejected H₁.



Documenting Results

When documenting the results of this study, the following textual summary and Table 6.9 would be appropriate to include:

In order to assess the most efficient instruction method to guide consumers in assembling the Able Chair, which requires one standard screwdriver, 105 adults were recruited and randomly assigned to assemble the chair independently using one of three types of instructions: (a) text only, (b) text with illustrations, or (c) an instructional video. The researcher unobtrusively recorded the duration of each

assembly (in minutes). All of the participants completed the assembly in under 100 minutes, which was our initial intent.

Those who used the text with illustrations ($\mu = 92.0$ minutes) completed the assembly significantly faster than those who used the text-only instructions ($\mu = 97.6$) ($p = .022$); additionally, those who were given the video ($\mu = 89.8$) also finished the assembly significantly faster than those who used the text-only instructions ($\mu = 97.6$) ($p = .001$).

Although, on average, those who used the video ($\mu = 89.8$) completed the job faster than those who used the text with illustrations ($\mu = 92.0$), this 2.2-minute difference is not statistically significant ($p = .536$) (see table below).

Results of ANOVA for *time*.

Groups	p
$\mu(\text{Text}) = 97.63 : \mu(\text{Text with illustrations}) = 92.00$.022*
$\mu(\text{Text}) = 97.63 : \mu(\text{Video}) = 80.63$.001*
$\mu(\text{Text with illustrations}) = 92.00 : \mu(\text{Video}) = 89.77$.536

*Statistically significant difference detected between groups ($p \leq .05$).



Occasionally, a p value may be close to .05 (e.g., $p = .066$). In such instances, you may feel compelled to comment that the p value of .066 is *approaching* statistical significance. While the optimism may be commendable, this is a common mistake. The term *approaching* wrongly implies that the p value is a dynamic variable—that it is in motion and on its way to crossing the .05 finish line, but this is not at all the case. The p value is a static variable, meaning that it is not in motion—the value of .066 is no more *approaching* .05 than it is *approaching* .07. Think of the p value as parked; it is not going anywhere, in the same way that a parked car is neither approaching nor departing from the car parked in front of it, no matter how close those cars are parked to each other. At best, one could state that the p value of .066 is *close* to the .05 α level and that it would be interesting to consider monitoring this variable should this experiment be repeated at some future point.

Here is a simpler way to think about this: $2 + 2 = 4$, and 4 is not approaching 3 or 5; it is just 4, and it is not drifting in any direction.

OVERVIEW—KRUSKAL-WALLIS TEST

One of the pretest criteria that must be met prior to running an ANOVA states that the data from each group must be normally distributed (Figure 6.9); minor variations in the normal distribution are acceptable. Occasionally, you may encounter data that are substantially skewed (Figure 6.10), bimodal (Figure 6.11), or flat (Figure 6.12) or that may

have some other atypical distribution. In such instances, the Kruskal-Wallis statistic is an appropriate alternative to the ANOVA test.

Figure 6.9 Normal distribution.

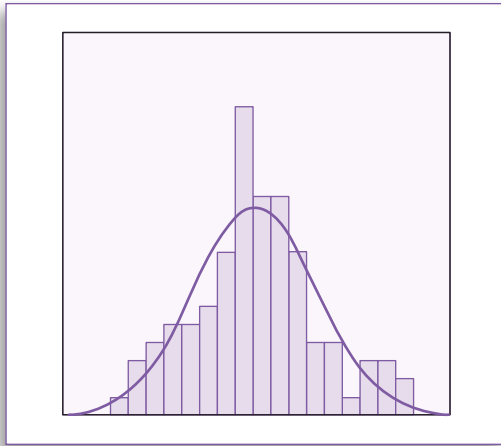


Figure 6.10 Skewed distribution.



Figure 6.11 Bimodal distribution.

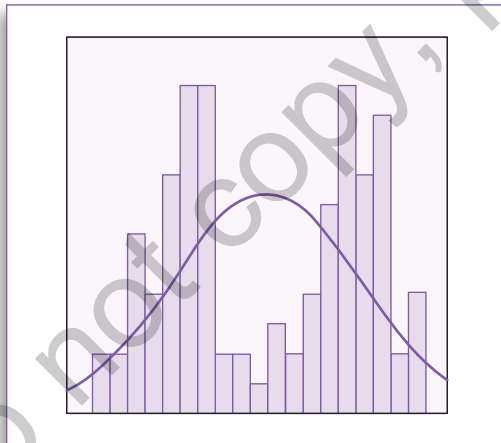
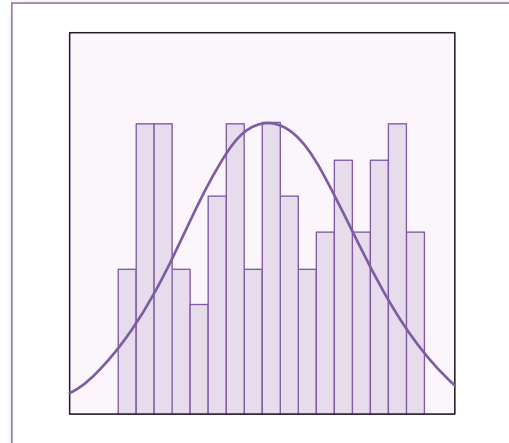


Figure 6.12 Flat distribution.



Test Run

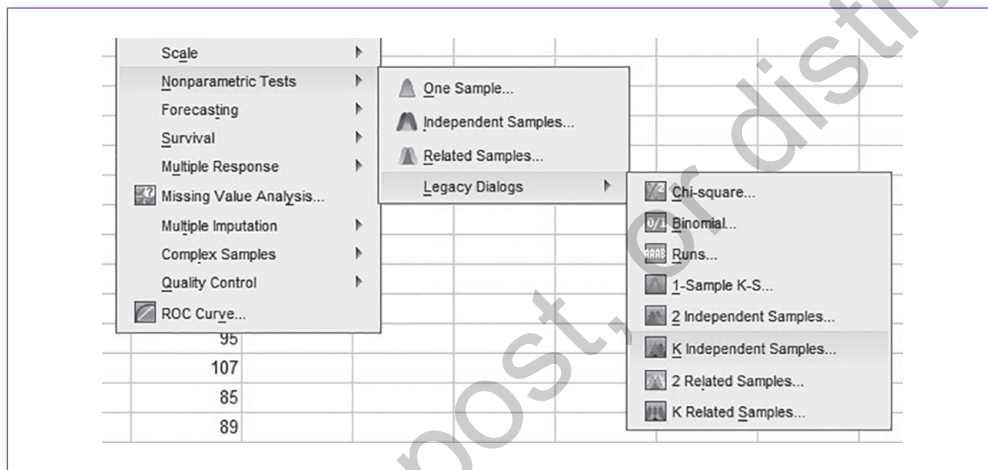
For exemplary purposes, we will run the Kruskal-Wallis test using the same data set (**Ch 06 - Example 01 - ANOVA and Kruskal-Wallis Test.sav**), even though the data are

normally distributed. This will enable us to compare the results of an ANOVA test with the results produced by the Kruskal-Wallis test.

1. On the main screen, click on *Analyze, Nonparametric Tests, Legacy Dialogs, K Independent Samples* (Figure 6.13).

Figure 6.13

Ordering the Kruskal-Wallis test: click on *Analyze, Nonparametric Tests, Legacy Dialogs, K Independent Samples*.



2. In the *Two-Independent Samples Tests* window, move *time* to the *Test Variable List* panel.
3. Move *group* to the *Grouping Variable* box (Figure 6.14).
4. Click on *group(??)*, then click on *Define Range*.
5. In the *Several Independent Samples: Define Range* window, for *Minimum*, enter 1; for *Maximum*, enter 3 (since the groups are numbered 1 [for Text] through 3 [for Video]) (Figure 6.15).
6. Click *Continue*; this will close this window.
7. In the *Tests for Several Independent Samples* window, click on *OK*.

Results

The Kruskal-Wallis result is found in the *Test Statistics* table (Table 6.10); the Asymp. Sig. statistic rendered a *p* value of .005; since this is less than α (.05), we would conclude that there is a statistically significant difference (somewhere) among the three types of assembly instructions, but we still need to conduct pairwise (post hoc type) analyses to determine which group(s) outperformed which.

Figure 6.14

In the *Tests for Several Independent Samples* window, move *time* to *Test Variable List* panel, and move *group* to the *Grouping Variable* box.

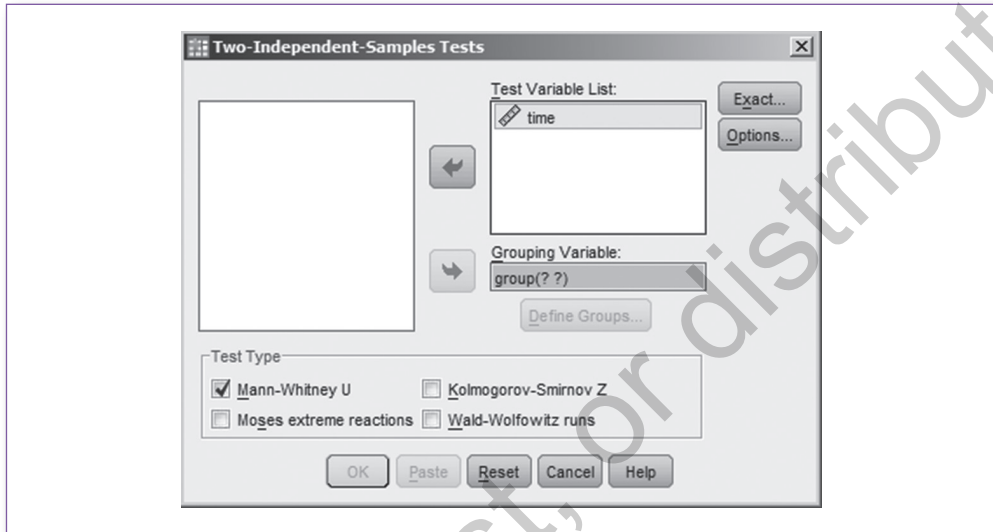


Figure 6.15

In the *Several Independent Samples: Define Range* window, for *Minimum*, enter 1; for *Maximum*, enter 3.

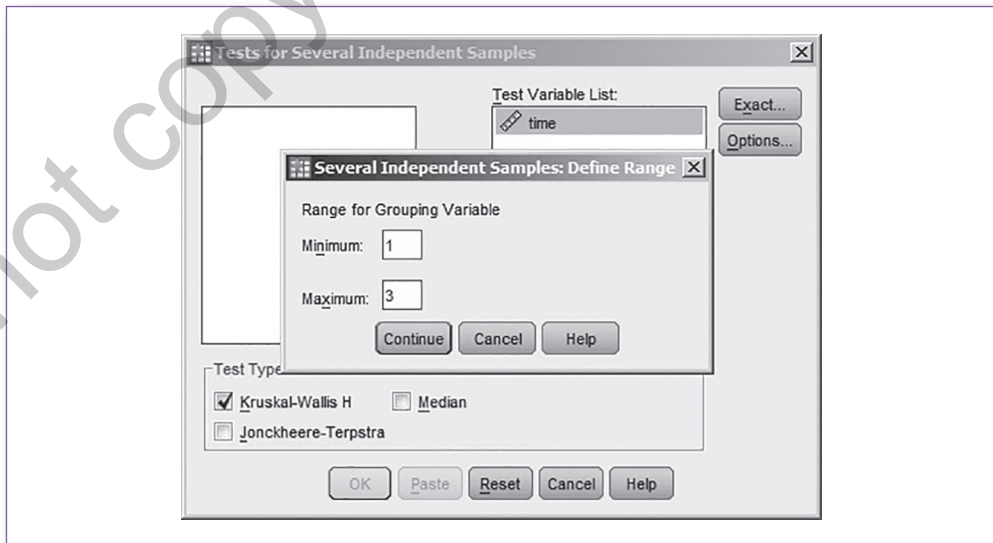


Table 6.10 Kruskal-Wallis p Value = .005

Test Statistics ^{a,b}	
	time
Chi-square	10.623
df	2
Asymp. Sig.	.005

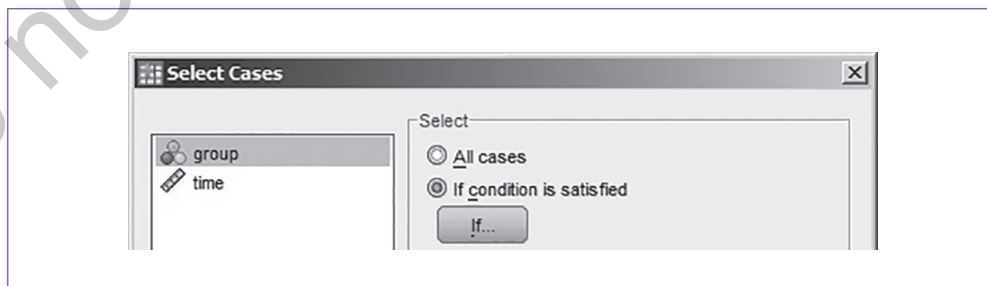
a. Kruskal Wallis Test
b. Grouping Variable: group

The ANOVA test provides a variety of post hoc options (e.g., Tukey, Sidak); although the Kruskal-Wallis test does not include a post hoc menu, we can take a few extra steps to process pairwise comparisons among the groups using the Kruskal-Wallis test. We will accomplish this using the *Select Cases* function to select two groups at a time and run separate Kruskal-Wallis tests for each pair. First, we will select and process Text : Text with illustrations, then Text : Video, and finally Text with illustrations : Video.

1. Click on the *Select Cases* icon.
2. In the *Select Cases* window, click on *If condition is satisfied* (Figure 6.16).

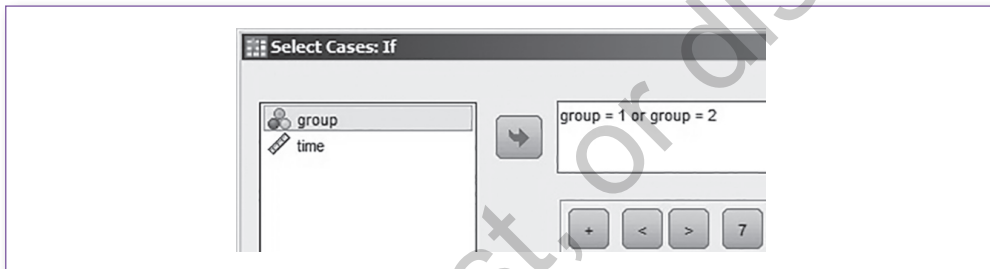
Figure 6.16

In the *Select Cases* window, click on *If condition is satisfied*, then click on *If*.



3. Click on *If*.
4. In the *Select Cases: If* window, specify the pair of groups that you want selected (Figure 6.17):
 - On the first pass through this process, enter **group = 1 or group = 2**.
 - On the second pass, enter **group = 1 or group = 3**.
 - On the third pass, enter **group = 2 or group = 3**.

Figure 6.17 In the *Select Cases: If* window, specify the pair of groups that you want selected.



5. Click *OK*.
6. Now that only two groups are selected, run the Kruskal-Wallis procedure from Step 1 and record the *p* value produced by each run; upon gathering these figures, you will be able to assemble a Kruskal-Wallis post hoc table (Table 6.11). NOTE: You can keep using the parameters specified from the previous run(s).

Table 6.11 Pairwise *p* Values for the Kruskal-Wallis Test (Manually Assembled)

Groups	<i>p</i>
Text : Text with illustrations	.020*
Text : Video	.002*
Text with illustrations : Video	.371

*Statistically significant difference detected between groups ($p \leq .05$).

To finalize this discussion, consider Table 6.12, which shows the *p* values produced by the ANOVA Tukey post hoc test compared alongside the *p* values produced by the Kruskal-Wallis test.

In addition to noting the differences in the pairwise p values (Table 6.12), remember that the ANOVA test produced an initial p value of .001 (which we read before the paired post hoc tests), whereas the Kruskal-Wallis produced an initial (overall) p value of .005. The difference in these p values is due to the internal transformations that the Kruskal-Wallis test conducts on the data. If one or more substantial violations are detected when running the pretest checklist for the ANOVA, the Kruskal-Wallis test is considered a viable alternative.

Table 6.12 ANOVA and Kruskal-Wallis Pairwise Post Hoc p values

Groups	ANOVA p	Kruskal-Wallis p
Text : Text with illustrations	.022*	.020*
Text : Video	.001*	.002*
Text with illustrations : Video	.536	.371

*Statistically significant difference detected between groups ($p \leq .05$).

GOOD COMMON SENSE

When carrying statistical results into the real world, there are some practical considerations to take into account. Using this example, suppose the manufacturer's goal was to have assembly instructions that would enable consumers to put the Able Chair together in under 100 minutes, in which case, any of the instructional methods would be appropriate. In this case, the researcher may opt to conduct a follow-up survey, asking consumers which form of assembly instructions they would prefer (text only, text with illustrations, or a video).

Cost-effectiveness may be another influential factor when it comes to applying these findings in the real world. Suppose it costs the same (20 cents) to provide text-only printed instructions or text with illustrations; clearly, text with illustrations would be the proper choice. Furthermore, suppose it costs \$1.00 per video. Considering that the video costs 5 times more than the written instructions, and the video saves only 2.2 minutes in assembly time (which was found to be statistically insignificant: $p = .536$, $\alpha = .05$), the video may not be the preferred choice. Additionally, if the consumer does not have access to a playback device, or one with a sufficiently large display, the video may not be useful.

The point is, statistical analysis can provide precise results that can be used in making (more) informed decisions, but in addition to statistical results, other factors may be considered when it comes to making decisions in the real world.

Another issue involves the capacity of the ANOVA model. Table 6.8 and the combinations formula (**unique pairs** = $G! \div (2 \times (G - 2)!)$) reveal that as more groups are included, the number of ANOVA post hoc paired comparisons increases substantially. A 5-group design would render 10 unique comparisons, 6 groups would render 15, and a 10-group design would render 45 unique comparisons, along with their corresponding p values. While SPSS or any statistical software would have no problem processing these figures,

there would be some real-world challenges to address. Consider the pretest criteria—in order for the results of an ANOVA test to be considered robust, there should be a minimum n of 30 per group. Hence, for a design involving 10 groups, this would require an overall n of at least 300. In addition, a 10-group study would render 45 unique pairwise comparisons in the ANOVA post hoc table, which, depending on the nature of the data, may be a bit unwieldy when it comes to interpretation and overall comprehension of the results.

Key Concepts

- ANOVA
- Pretest checklist
 - Normality
 - Homogeneity of variance
 - n
- Post hoc tests
 - Tukey
 - Sidak
- Hypothesis resolution
- Documenting results
- Kruskal-Wallis test
- Good common sense

Practice Exercises

Use the prepared SPSS data sets (download from study.sagepub.com/knappstats2e).

NOTE: These practice exercises and data sets are the same as those in Chapter 5 (“ t Test and Mann-Whitney U Test”), except that instead of the two-group designs, additional data have been included to facilitate ANOVA processing: Exercises 6.1 through 6.8 have three groups, and Exercises 6.9 and 6.10 have four groups.

Exercise 6.1

You want to determine if meditation can reduce resting pulse rate. Participants were recruited and randomly assigned to one of three groups: Members of Group 1 (the control group) will not meditate; members of Group 2 (the first treatment group) will meditate for 30 minutes per day on Mondays, Wednesdays, and Fridays over the course of 2 weeks; and members of Group 3 (the second treatment group) will meditate for 30 minutes a day 6 days a week, Monday through Saturday. At the end of the study, you gathered the resting pulse rates for each participant.

Data set: **Ch 06 - Exercise 01A.sav**

Codebook

Variable: group

Definition: Group number

Type: Categorical (1 = No meditation, 2 = Meditates 3 days, 3 = Meditates 6 days)

Variable: pulse

Definition: Pulse rate (beats per minute)

Type: Continuous

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 06 - Exercise 01B.sav**.

Exercise 6.2

You want to determine if pairing an incoming freshman with a sophomore in a mentor-mentee relationship will enhance the freshman's overall grade. You recruit sophomores who are willing to mentor students in their majors for their first terms. You then recruit freshmen who are interested in having mentors. Freshmen who apply to this program will be sequentially assigned to one of three groups: Group 1 will be the control group (no mentor), Group 2 will be the in-person mentor group, and Group 3 will be the e-mentor group. Those in the in-person mentor group are to meet in person once a week at a time of their choosing, and those in the e-mentor group will communicate digitally at least once a week. All freshmen, in each group, agree to submit their transcripts at the conclusion of the term.

Data set: **Ch 06 - Exercise 02A.sav**

Codebook

Variable: group

Definition: Mentor group assignment

Type: Categorical (1 = No mentor, 2 = In-person mentor, 3 = E-mentor)

Variable: grade
 Definition: Student's overall grade
 Type: Continuous (0–100)

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 06 - Exercise 02B.sav**.

Exercise 6.3

Clinicians at a nursing home facility want to see if giving residents plants to tend to will help lower depression. To test this idea, each resident is randomly assigned to one of three groups: Group 1, Control group (no plant); Group 2, Bamboo; or Group 3, Cactus. Each plant comes with a card detailing care instructions. After 90 days, all participants will complete the Acme Depression Scale, a 10-question instrument that renders a score between 1 and 100 (1 = low depression, 100 = high depression).

Data set: **Ch 06 - Exercise 03A.sav**

Codebook

Variable: group
 Definition: Plant group assignment
 Type: Categorical (1 = No plant, 2 = Bamboo, 3 = Cactus)

Variable: mood
 Definition: Depression level
 Type: Continuous (1 = low depression, 100 = high depression)

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).

- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 06 - Exercise 03B.sav**.

Exercise 6.4

You want to determine if chocolate enhances mood. Subjects will be recruited and randomly assigned to one of three groups: Those in Group 1 will be the control group and will eat their regular diet. Those in Group 2 will eat their usual meals and have one piece of chocolate at breakfast, lunch, and dinner over the course of a week. Those in Group 3 will eat their meals as usual and have two pieces of chocolate at breakfast, lunch, and dinner over the course of a week. At the end of the week, all participants will complete the Acme Mood Scale (1 = extremely bad mood, 100 = extremely good mood).

Data set: **Ch 06 - Exercise 04A.sav**

Codebook

Variable: group

Definition: Chocolate dosage group assignment

Type: Categorical (1 = No chocolate, 2 = Chocolate [1 per meal],
3 = Chocolate [2 per meal])

Variable: Mood

Definition: Acme Mood Scale score

Type: Continuous (1 = extremely bad mood, 100 = extremely good mood)

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 06 - Exercise 04B.sav**.

Exercise 6.5

During flu season, the administrators at a walk-in health clinic want to determine if providing patients with a pamphlet or a video will increase their receptivity to flu shots.

Each patient will be given a ticket at the check-in desk with a 1, 2, or 3 on it; the tickets will be issued in (repeating) sequence. Once escorted to the exam room, patients with number 1 tickets will serve as control participants and will not be offered any flu shot informational material. Patients with number 2 tickets will be given a flu shot information pamphlet. Patients with number 3 tickets will be shown a brief video covering the same information as contained in the pamphlet. At the end of the day, the charts will be reviewed and three entries made in the database: total number of flu shots given to patients in Group 1, total number of flu shots given to patients in Group 2, and total number of flu shots given to patients in Group 3.

Data set: **Ch 06 - Exercise 5A.sav**

Codebook

Variable: group
 Definition: Flu media group assignment
 Type: Categorical (1 = Nothing, 2 = Flu shot pamphlet, 3 = Flu shot video)

Variable: shots
 Definition: Total number of flu shots given in a day (for each group)
 Type: Continuous

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 06 - Exercise 5B.sav**.

Exercise 6.6

You want to determine if watching a video of a comedy with a laugh track enhances enjoyment. Subjects will be recruited and randomly assigned to one of three groups: Those in Group 1 (the control group) will watch the video without the laugh track, those assigned to Group 2 will watch the same video with the sound of a 50-person audience included in the soundtrack, and those assigned to Group 3 will watch the same video with the sound of a 100-person audience included in the soundtrack. Each participant will watch the video individually; no others will be present in the room. Immediately

following the video, each participant will be asked to rate how enjoyable the show was on a scale of 1 to 5 (1 = not very enjoyable, 5 = very enjoyable).

Data set: **Ch 06 - Exercise 06A.sav**

Codebook

Variable:	group
Definition:	Laugh track group assignment
Type:	Categorical (1 = No laugh track, 2 = Laugh track at 50, 3 = Laugh track at 100)
Variable:	enjoy
Definition:	Enjoyment of video
Type:	Continuous (1 = not very enjoyable, 5 = enjoyable)

- Write the hypotheses.
- Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 06 - Exercise 06B.sav**.

Exercise 6.7

In an effort to determine the effectiveness of light therapy to alleviate depression, you recruit a group of subjects who have been diagnosed with depression. The subjects are randomly assigned to one of three groups: Group 1 will be the control group—members of this group will receive no light therapy. Members of Group 2 will get light therapy for 1 hour on even-numbered days over the course of 1 month. Members of Group 3 will get light therapy every day for 1 hour over the course of 1 month. After 1 month, all participants will complete the Acme Mood Scale, consisting of 10 questions; this instrument renders a score between 1 and 100 (1 = extremely bad mood, 100 = extremely good mood).

Data set: **Ch 06 - Exercise 07A.sav**

Codebook

Variable:	group
Definition:	Light therapy group assignment

Type: Categorical (1 = No light therapy, 2 = Light therapy: even days, 3 = Light therapy: every day)

Variable: mood

Definition: Acme Mood Scale score

Type: Continuous (1 = extremely bad mood, 100 = extremely good mood)

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using **Ch 06 - Exercise 07B.sav**.

Exercise 6.8

It is thought that exercising early in the morning will provide better energy throughout the day. To test this idea, subjects are recruited and randomly assigned to one of three groups: Members of Group 1 will constitute the control group and not be assigned any walking. Members of Group 2 will walk from 7:00 to 7:30 a.m., Monday through Friday, over the course of 30 days. Members of Group 3 will walk from 7:00 to 8:00 a.m., Monday through Friday, over the course of 30 days. At the conclusion of the study, each participant will answer the 10 questions on the Acme End-of-the-Day Energy Scale. This instrument produces a score between 1 and 100 (1 = extremely low energy, 100 = extremely high energy).

Data set: **Ch 06 - Exercise 08A.sav**

Codebook

Variable: group

Definition: Walking group assignment

Type: Categorical (1 = No walking, 2 = Walking: 30 Minutes, 3 = Walking: 60 minutes)

Variable: mood

Definition: Energy level

Type: Continuous (1 = extremely low energy, 100 = extremely high energy)

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using **Ch 06 - Exercise 08B.sav**.

NOTE: Exercises 9 and 10 involve four groups each.

Exercise 6.9

The Acme Company claims that its new reading lamp increases reading speed; you want to test this. You will record how long (in seconds) it takes for participants to read a 1,000-word essay. Participants will be randomly assigned to one of four groups: Group 1 will be the control group; they will read the essay using regular room lighting. Those in Group 2 will read the essay using the Acme lamp. Those in Group 3 will read the essay using a generic reading lamp. Those in Group 4 will read the essay using a flashlight.

Data set: **Ch 06 - Exercise 09A.sav**

Codebook

Variable: group
 Definition: Lighting group assignment
 Type: Categorical (1 = Room lighting, 2 = Acme lamp, 3 = Generic lamp, 4 = Flashlight)

Variable: Seconds
 Definition: The time it takes to read the essay
 Type: Continuous

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 06 - Exercise 09B.sav**.

Exercise 6.10

You want to find out if music enhances problem solving. Subjects will be recruited and randomly assigned to one of four groups: Those in Group 1 will serve as the control group and will be given a standard 100-piece jigsaw puzzle to solve in a quiet room. Participants in Group 2 will be given the same puzzle to assemble, but instead of silence, there will be classical music playing at a soft volume (30 decibels [dB]) in the room. Participants in Group 3 will be given the same puzzle to assemble using the same classical music, but the music will be played at a moderate volume (60 dB). Participants in Group 4 will be given the same puzzle to assemble using the same classical music, but the music will be played at a loud volume (90 dB). You will record the time (in seconds) that it takes for each person to complete the puzzle.

Data set: **Ch 06 - Exercise 10A.sav**

Codebook

Variable: group
 Definition: Music group number
 Type: Categorical (1 = No music, 2 = Music at 30 dB, 3 = Music at 60 dB, 4 = Music at 90 dB)

Variable: seconds
 Definition: Time required to complete the puzzle
 Type: Continuous

- a. Write the hypotheses.
- b. Run each criterion of the pretest checklist (normality, homogeneity of variance, and n) and discuss your findings.
- c. Run the ANOVA test and document your findings (ns , means, and significance [p value], hypothesis resolution).
- d. Write an abstract up to 200 words detailing a summary of the study, the ANOVA test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using **Ch 06 - Exercise 10B.sav**.