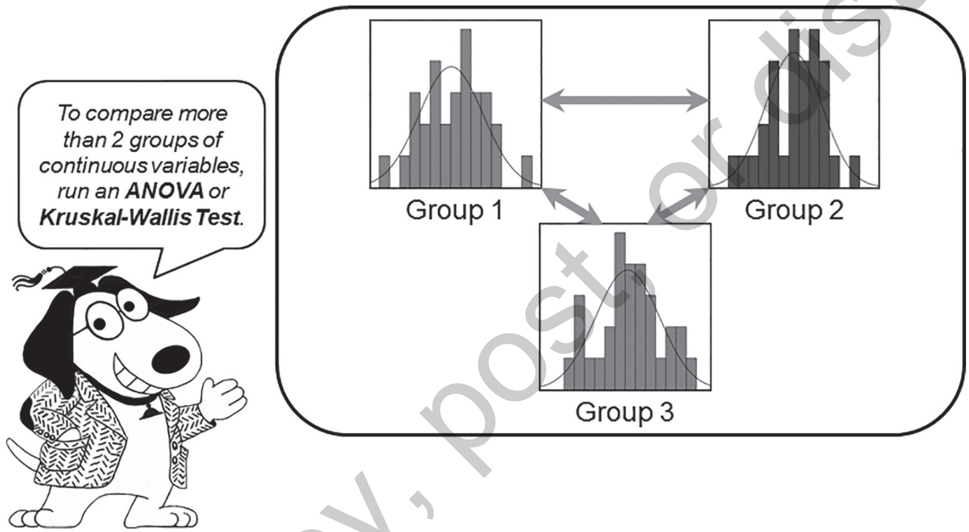
CHAPTER 5

ANOVA and Kruskal-Wallis Test



Three is a magic number.

—Bob Dorough

Learning Objectives

Upon completing this chapter, you will be able to do the following:

- 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.

(Continued)

(Continued)

- Calculate the unique pairs formula.
- Resolve the hypotheses.
- Know when and how to run and interpret the Kruskal-Wallis test.
- Write an appropriate abstract.

WHEN TO USE THIS STATISTIC

Guidelines for ANOVA and Kruskal-Wallis Tests

Overview: This statistic is for designs that involve more than two groups to determine which group(s) (if any) outperformed another.

Variables: This statistic requires two variables for each record: (1) a categorical variable to designate the group, and (2) a continuous variable to contain the outcome score.

Results: Among those diagnosed with anxiety, we separated our participants into three groups and measured their pulse rates after 30 minutes of treatment. The mean pulse rates were 84.1 for the control group, 79.9 for the pet therapy group, and 80.1 for the meditation group. The pulse rate for the pet therapy group is statistically significantly lower than the control group (p = .025, α = .05).

VIDEOS

The videos for this chapter are **Ch 05 – ANOVA.mp4** and **Ch 05 – Kruskal-Wallis Test .mp4.** These videos provide overviews of these tests, instructions for carrying out the pretest checklist, run, and interpreting the results of each test using the data set: **Ch 05 – 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 4 (*t* 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 4, which involved measuring the pulse rate of anxious participants from two groups: Group 1 (Control), Group 2 (Pet therapy), and now a third group: Group 3 (Meditation). The ANOVA test will reveal which (if any) of these three treatments statistically significantly outperforms the others in terms of lowering the resting pulse rate.

OVERVIEW—ANOVA

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

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 alternate test is explained near the end of this chapter.

Example

A research team has recruited a group of individuals who have been diagnosed with acute stress disorder to determine the effectiveness of supplemental nonpharmaceutical treatments for reducing stress: (1) no supplemental therapy, (2) pet therapy with certified therapy dogs, or (3) meditation.

Research Question

Is pet therapy or meditation effective in reducing stress among those diagnosed with acute stress disorder?

Groups

A researcher recruits a total of 90 participants who meet the diagnostic criteria for acute stress disorder. 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 on a sequential basis (first assigned to Group 1, second assigned to Group 2, third assigned to Group 3, fourth assigned back to Group 1, etc.). Those assigned to Group 1 will constitute the control group and will be instructed to sit and relax for 30 minutes (with no treatment). Those in Group 2 will receive 30 minutes of pet therapy, and those in Group 3 will meditate for 30 minutes.

Procedure

Each participant will be guided to a room with a comfortable sofa. Those in the control group will be instructed to just sit and relax. Those in the pet therapy group will be introduced to the therapy dog by name and instructed that the participant may hold the dog in his or her lap, pet the dog on the sofa, brush the dog, or give the dog the allotted treats in whatever combination they wish. Those in the meditation group will listen to a recording of gentle music with a narrative taking the participant through a guided meditation. After 30 minutes, the researcher will return to each participant to measure the participant's pulse rate and dismiss him or her. The lower pulse rate would reflect more relaxation.

Hypotheses

The null hypothesis (H_0) is phrased to anticipate that the treatments (pet therapy and meditation) fail to reduce anxiety (as measured by pulse rate), indicating that on average, participants in each of the three groups all have about the same pulse rates; in other words, no group outperforms any other when it comes to lowering the participant's pulse rates. 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 pulse rates across the groups.

H_1: There is a difference in pulse rates across the groups.

Admittedly, H_1 is phrased fairly broadly. The *Post Hoc Multiple Comparisons* table, which is covered in the Results section, will identify which treatment(s), if any, outperformed which.

Data Set

Use the following data set: **Ch 05 – Example 01 – ANOVA and Kruskal-Wallis Test.sav**.

Notice that this data set has 90 records; the first 60 records (rows) are the same as the *t* test and Mann-Whitney *U* test example data set used in Chapter 4 (records 61 through 90 are new for Group 3):

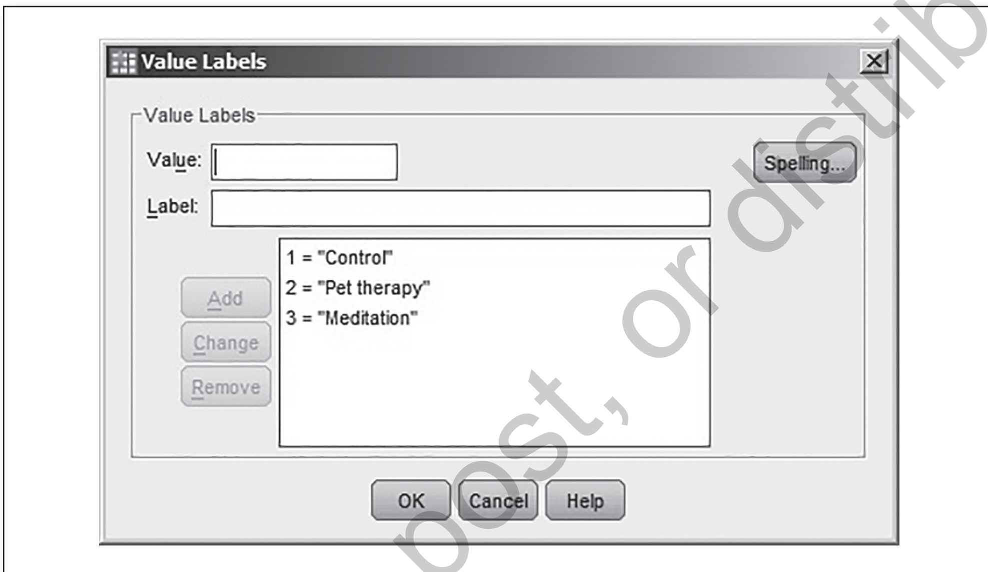
Codebook

Variable:	Group
Definition:	Group assignment
Type:	Categorical (1 = Control, 2 = Pet therapy, 3 = Meditation)
Variable:	Pulse
Definition:	Heartbeats per minute measured 30 minutes after start
Type:	Continuous

NOTE: In this data set, records (rows) 1 through 30 are for Group 1 (Control), records 31 through 60 are for Group 2 (Pet therapy), and records 61 through 90 are for Group 3 (Meditation). 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 *Values* menu for the variable *Group*, you will see that the label *Meditation* for the third group has been assigned to the value 3 (Figure 5.1).

Figure 5.1 Value labels for a three-group ANOVA analysis.



Pretest Checklist

ANOVA Pretest Checklist

☑ 1. Normality*

☑ 2. *n* quota*

☑ 3. Homogeneity of Variance **

*Run prior to ANOVA test

**Results produced upon ANOVA test run

NOTE: If any of the pretest checklist criteria are not satisfied, rerun the analysis using the nonparametric version of this test: the Kruskal-Wallis test (p. 125).

The statistical pretest checklist for the ANOVA is similar to the *t* test: (1) normality, (2) *n*, and (3) 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 *Control* group (*Group = 1*); the selection criteria would be *Group = 1*. Next, run a histogram (with normal curve) on the variable *Pulse*. For more details on this procedure, refer to **Chapter 3: Descriptive Statistics**, and the following section: SPSS—Descriptive Statistics: Continuous Variable (Age) Select by Categorical Variable (Gender)—Females Only; see the star (★) icon on page 62.

Then repeat the process for the *Pet therapy* group (*Group* = 2), and finally, repeat the process a third time for the *Meditation* group (*Group* = 3).

This will produce three histograms with normal curves—one for the scores in the *Control* group, a second for the scores in the *Pet therapy* group, and a third for the *Meditation* group. The histograms should resemble the graphs shown in Figures 5.2, 5.3, and 5.4.

Figure 5.2 Histogram of *Pulse* for Group 1: *Control*.

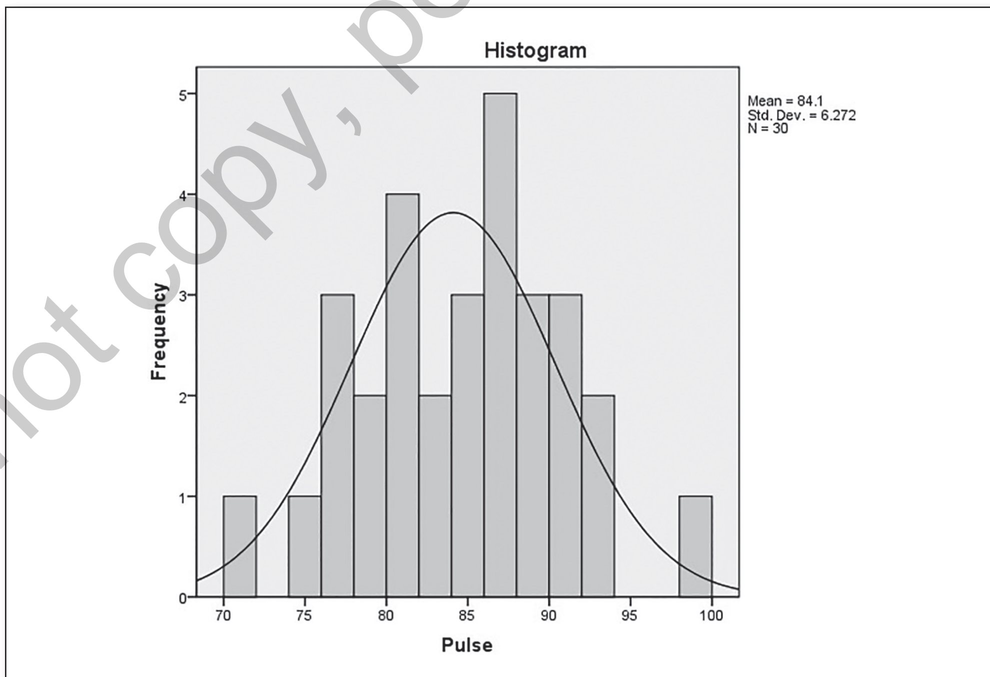


Figure 5.3 Histogram of *Pulse* for Group 2: *Pet therapy*.

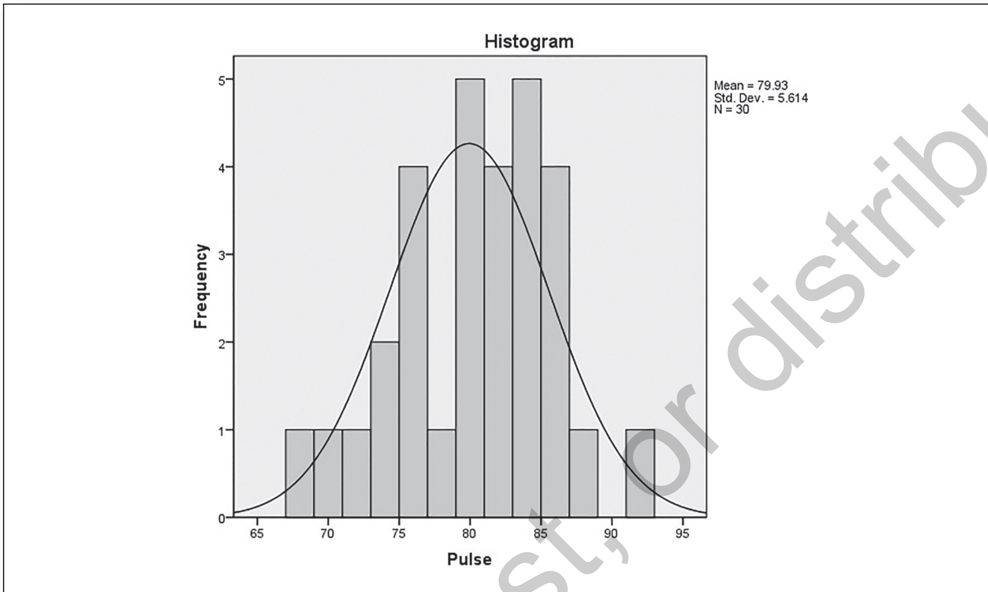
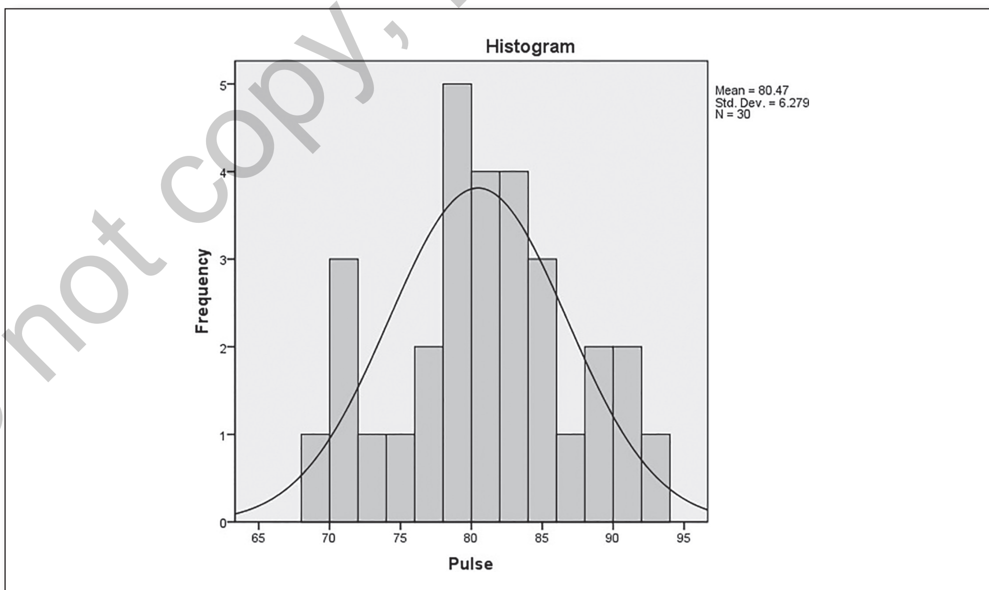


Figure 5.4 Histogram of *Pulse* for Group 3: *Meditation*.



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 **Chapter 3: Descriptive statistics**, and the following section: SPSS—(Re)Selecting All Variables; see the star (★) icon on page 68.

Pretest Checklist Criterion 2—*n* Quota

As with the *t* test, technically, you can run an ANOVA test with an *n* of any size in each group, but when *n* is at least 30 in each group, the ANOVA is considered more robust. The histograms with normal curves (Figures 5.2, 5.3, and 5.4) include *n*s associated with each group; in this case $n = 30$ for each group. Normality is considered a more important criterion than the *n* quota; if a group is showing a normal distribution but the n is low, you can proceed with confidence.

Pretest Checklist Criterion 3—Homogeneity of Variance

Homogeneity pertains to sameness. The homogeneity of variance criterion involves checking that the variances among the 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 groups. In this case, the variance for *Pulse* in the *Control* group is 39.3 (derived from Figure 5.2: $6.272^2 = 39.338$), the variance for *Pulse* in the *Pet therapy* group is 31.7 (derived from Figure 5.3: $5.614^2 = 31.517$), and the variance for *Pulse* in the *Meditation* group is 39.4 (derived from Figure 5.4: $6.279^2 = 39.426$). When looking at the *Pulse* variances from these three groups (39.3, 31.7, and 39.4), clearly none of these figures are more than twice any of the others, so we would expect that the homogeneity of variance test would pass.

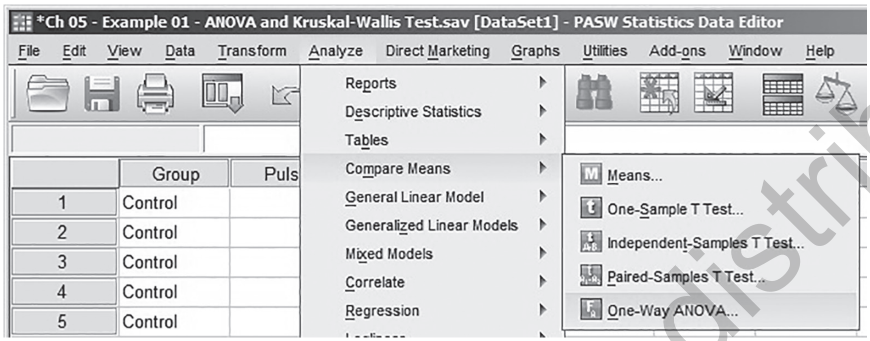
The *homogeneity of variance test* is an option selected during the ANOVA run. If the homogeneity of variance test renders a significance (*p*) value that is greater than .05, then this suggests that there are no statistically significant differences among the variances of the groups. This would mean that the data pass the homogeneity of variance test.

Test Run

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

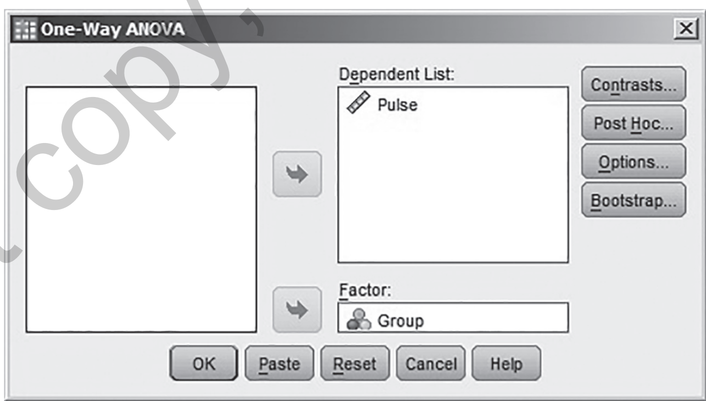
1. First, be sure to reactivate all of the records; the easiest way to do this is to delete the *filter_$* variable, or click on the *Select Cases* icon and under *Select*, click on *All cases*.
2. On the main screen, click on *Analyze, Compare Means, One-Way ANOVA* (Figure 5.5).

Figure 5.5 Running an ANOVA test.



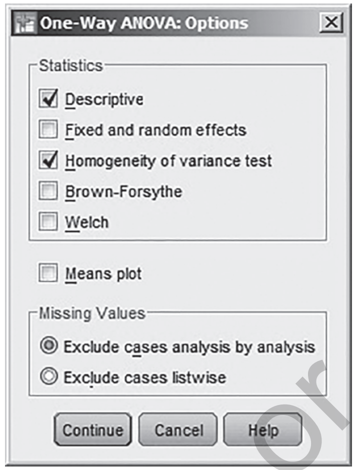
3. On the *One-Way ANOVA* menu, move the continuous variable that you wish to analyze (*Pulse*) into the *Dependent List* window, and move the variable that contains the categorical variable that specifies the group (*Group*) into the *Factor* window (Figure 5.6).

Figure 5.6 The *One-Way ANOVA* menu.



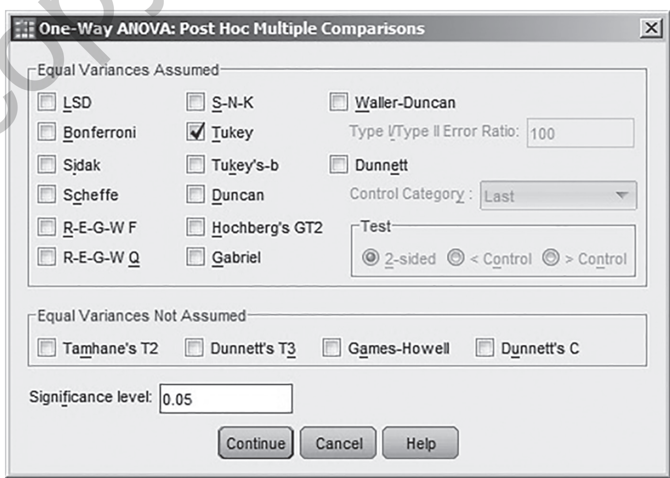
4. Click on the *Options* button. On the *One-Way ANOVA: Options* menu, check *Descriptive* and *Homogeneity of variance test;* then click on the *Continue* button (Figure 5.7). This will take you back to the *One-Way ANOVA* menu.

Figure 5.7 The *One-Way ANOVA: Options* menu.



5. Click on the *Post Hoc* button.
6. This will take you to the *One-Way ANOVA: Post Hoc Multiple Comparisons* menu (Figure 5.8).

Figure 5.8 The *One-Way ANOVA: Post Hoc Multiple Comparisons* menu.



7. If you were to run the ANOVA test without selecting a post hoc test, then all it would return is a single p value; if that p is statistically significant, then 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 values for each pair of comparisons. This will become clearer in the Results section when we read the post hoc multiple comparisons table.

8. On the *One-Way ANOVA* menu (Figure 5.6), click on the *OK* button, and the ANOVA test will process.

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 ns*; in this case, each group has an n of 30, so check the *Tukey* checkbox; then click on the *Continue* button (this will take you back to the *One-Way ANOVA* menu, see Figure 5.6). If the groups had *different ns* (e.g., n(Group 1) = 40, n(Group 2) = 55, n(Group 3) = 36), then the **Sidak** post hoc test would be appropriate. If you do not know the *n*s for each group in advance, then just select either *Tukey* or *Sidak* and observe the *n*s on the resulting report; if you chose wrong, then go back and rerun the analysis using the appropriate post hoc test.

ANOVA Post Hoc Summary

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

Results

Pretest Checklist Criterion 3—Homogeneity of Variance

For the final item on the pretest checklist, Table 5.1 shows that the homogeneity of variance test produced a significance (p) value of .689. Because this is greater than the α level of .05, this tells us that there are *no statistically significant differences among the variances of the* Pulse *variable for the three groups analyzed*. In other words, the variances for *Pulse* are similar enough among the three groups: *Control*, *Pet therapy*, and *Meditation*; hence, we would conclude that the criteria of the homogeneity of variance has been satisfied.

Table 5.1 Homogeneity of variance test results.

Test of Homogeneity of Variances

Pulse

Levene Statistic	df1	df2	Sig.
.374	2	87	.689

Table 5.2 ANOVA test results comparing *Pulse* of *Control*, *Pet therapy*, and *Meditation*.

ANOVA

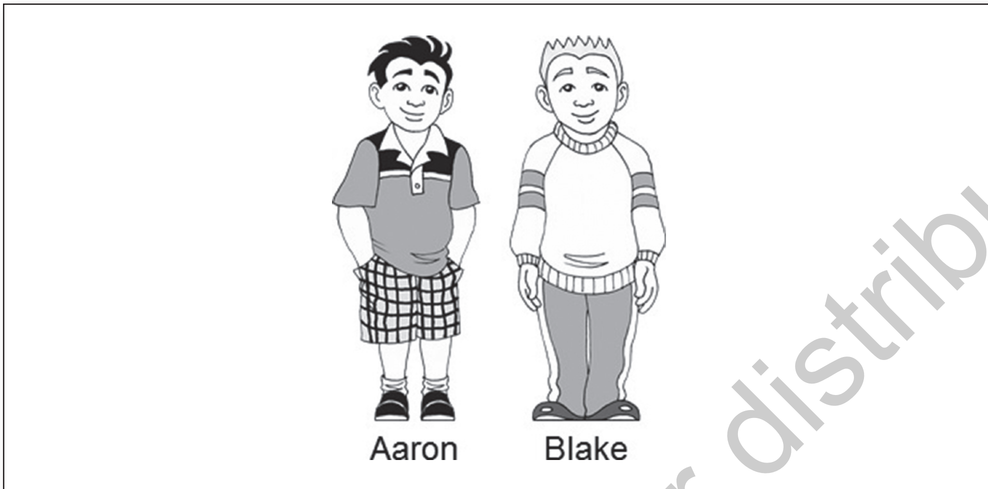
Pulse

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	308.467	2	154.233	4.196	.018
Within Groups	3198.033	87	36.759		
Total	3506.500	89			

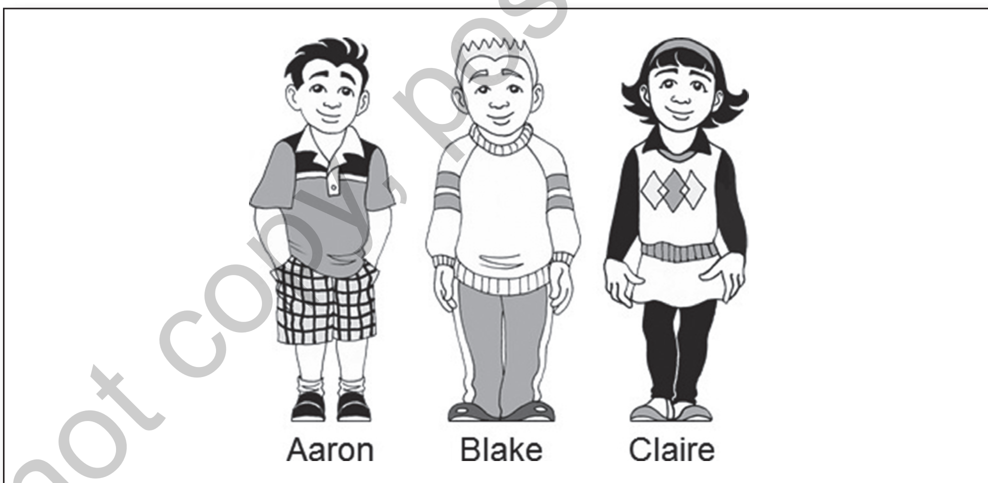
Next, we look at the ANOVA table (Table 5.2) and find a significance (p) value of .018. This is less than the α level of .05, so this tells us that there is a statistically significant difference somewhere among the (three) group means for *Pulse,* but unlike reading the results of the t test, we are not done yet.

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 is $\leq$ .05, there is no question as to which group is different from which—clearly, the mean from Group 1 is statistically significantly different from the mean of Group 2, but 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.

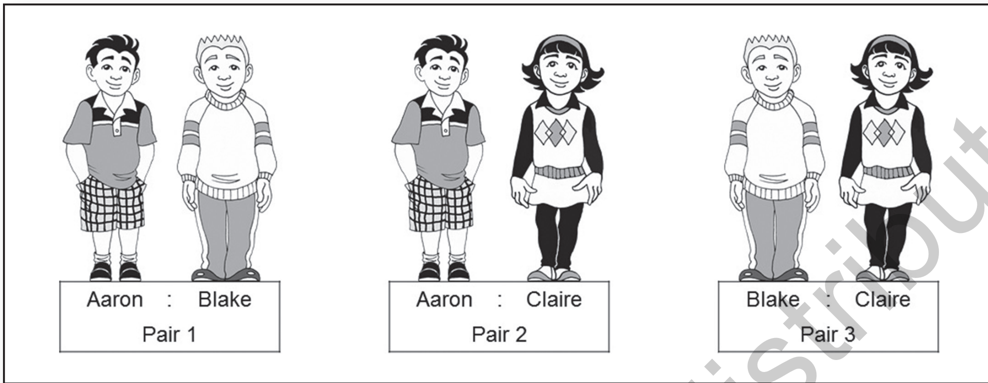
Consider this: Suppose you have *two* kids, Aaron and Blake. You are in the living room, and someone calls out from the den, "The kids are fighting again!" Because there are only *two* kids, you immediately know that the fight is between Aaron and Blake. This is akin to the t test, which involves comparing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



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



This time when someone 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 5.2) produced a statistically significant *p* value (Sig. = .018), 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 5.3). As with the three kids fighting, in this three-group design, there are three possible pairs of comparisons that we can assess in terms of (mean) *Pulse* for the groups. NOTE: Mean scores are drawn from the histograms with normal curves (Figures 5.2, 5.3, and 5.4); these figures are also on the *Descriptives* table produced in the ANOVA output report (not shown here).

Group 1 : Group 2	Group 1 : Group 3	Group 2 : Group 3
84.10 : 79.93	84.10 : 80.47	79.93 : 80.47
Pair 1	**Pair 2**	**Pair 3**

Pairs of means shown for Group 1—Control, Group 2—Pet therapy, and Group 3—Meditation

We will use means shown previously and Table 5.3 (*Multiple Comparisons*) to analyze the ANOVA test results. To summarize, the mean *Pulse* for each of the three groups: *Control* (M = 84.10), *Pet therapy* (M = 79.93), and *Meditation* (M = 80.47). We will assess each of the three pairwise score comparisons separately.

Comparison 1—*Control* : *Pet Therapy*

Table 5.3 compares the mean *Pulse* for the *Control* group with the mean *Pulse* rate for the *Pet therapy* group, which produces a Sig.(nificance) (*p*) of .025. Because the *p* is less than the .05 α level, this tells us that for *Pulse,* there is a statistically significant difference between *Control* (M = 84.10) and *Pet therapy* (M = 79.93).

Table 5.3 *ANOVA Post Hoc Multiple Comparisons* table shows a statistically significant difference between *Control* and *Pet therapy* (p = .025).

Multiple Comparisons

Pulse
Tukey HSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Control	Pet therapy	4.167*	1.565	.025	.43	7.90
	Meditation	3.633	1.565	.058	-.10	7.37
Pet therapy	Control	-4.167*	1.565	.025	-7.90	-.43
	Meditation	-.533	1.565	.938	-4.27	3.20
Meditation	Control	-3.633	1.565	.058	-7.37	.10
	Pet therapy	.533	1.565	.938	-3.20	4.27

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

Comparison 2—*Control* : *Meditation*

The second comparison in Table 5.4 is between *Control* and *Meditation*, which produces a Sig.(nificance) (p) of .058. Because the p is greater than the .05 α level, this tells us that for *Pulse*, there is no statistically significant difference between *Control* (M = 84.10) and *Meditation* (M = 80.47).

Table 5.4 *ANOVA Post Hoc Multiple Comparisons* table shows a statistically insignificant difference between *Control* and *Meditation* (p = .058).

Multiple Comparisons

Pulse
Tukey HSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Control	Pet therapy	4.167*	1.565	.025	.43	7.90
	Meditation	3.633	1.565	.058	-.10	7.37
Pet therapy	Control	-4.167*	1.565	.025	-7.90	-.43
	Meditation	-.533	1.565	.938	-4.27	3.20
Meditation	Control	-3.633	1.565	.058	-7.37	.10
	Pet therapy	.533	1.565	.938	-3.20	4.27

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

Comparison 3—*Pet Therapy* : *Meditation*

The third comparison in Table 5.5 is between *Pet therapy* and *Meditation*, which produces a Sig.(nificance) (p) of .938. Since the p is greater than the .05 α level, this tells us that for *Pulse,* there is no statistically significant difference between *Pet therapy* (M = 79.93) and *Meditation* (M = 80.47).

Table 5.5 *ANOVA Post Hoc Multiple Comparisons* table shows no statistically significant difference between *Pet therapy* and *Meditation* (p = .938).

Multiple Comparisons

Pulse
Tukey HSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Control	Pet therapy	4.167*	1.565	.025	.43	7.90
	Meditation	3.633	1.565	.058	-.10	7.37
Pet therapy	Control	-4.167*	1.565	.025	-7.90	-.43
	Meditation	-.533	1.565	.938	-4.27	3.20
Meditation	Control	-3.633	1.565	.058	-7.37	.10
	Pet therapy	.533	1.565	.938	-3.20	4.27

*. 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 that you would expect to see in a typical two-dimensional table. For example, in a multiplication table, you would see two 32s in the table because 4 × 8 = 32 and 8 × 4 = 32. Similarly, the Sig. column of the *Multiple Comparisons* table (Table 5.6) contains two p values of .025: one comparing *Control* to *Pet therapy* and the other comparing *Pet therapy* to *Control*. In addition, there are two .058 p values (*Control* : *Meditation* and *Meditation* : *Control*) and two .938 p values (*Pet therapy* : *Meditation* and *Meditation* : *Pet therapy*).

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 5.7).

Table 5.6 *ANOVA Post Hoc Multiple Comparisons* table containing double-redundant Sig. (*p*) values: *Control* : *Pet therapy* produces the same *p* value as *Pet therapy* : *Control* ($p = .025$).

Multiple Comparisons

Pulse
Tukey HSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Control	Pet therapy	4.167*	1.565	.025	.43	7.90
	Meditation	3.633	1.565	.058	-.10	7.37
Pet therapy	Control	-4.167*	1.565	.025	-7.90	-.43
	Meditation	-.533	1.565	.938	-4.27	3.20
Meditation	Control	-3.633	1.565	.058	-7.37	.10
	Pet therapy	.533	1.565	.938	-3.20	4.27

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

Table 5.7 Increasing groups substantially increases ANOVA post hoc multiple comparisons.

2 Groups Renders 1 Comparison	*3 Groups Renders 3 Comparisons*	*4 Groups Renders 6 Comparisons*
$G_1:G_2$	$G_1:G_2$	$G_1:G_2$
	$G_2:G_3$	$G_2:G_3$
	$G_1:G_3$	$G_3:G_4$
		$G_1:G_3$
		$G_2:G_4$
		$G_1:G_4$

NOTE: G = group.

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

Unique Pairs Formula

G = Number of groups

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

The preceding 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 × 2 × 3, which equals 6.

H_0 Hypothesis Resolution

To clarify the hypothesis resolution process, it is helpful to organize the findings in a table and use an asterisk to flag statistically significant difference(s) (Table 5.8).

NOTE: SPSS does not generate this table (Table 5.8) directly; you can assemble this table by gathering the means from the *Descriptives* table or the histograms with normal curves (Figures 5.2, 5.3, and 5.4) and the *p* values from the Sig. column in the *Multiple Comparisons* table (Table 5.3).

With this results table assembled, we can now revisit and resolve our pending hypotheses, which focuses on identifying the best way to reduce anxiety. To finalize this process, we will assess each hypothesis per the statistics contained in Table 5.8.

REJECT: H_0: There is no difference in pulse rates across the groups.

ACCEPT: H_1: There is a difference in pulse rates across the groups.

Table 5.8 Results summary of ANOVA for pulse.

Groups	*p*
Control (M = 84.10) : Pet therapy (M = 79.93)	.025*
Control (M = 84.10) : Mediation (M = 80.47)	.058
Pet therapy (M = 79.93) : Mediation (M = 80.47)	.938

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

Because we discovered a statistically significant difference among at least one pair of the relaxation techniques, we reject H_0 and accept H_1. Specifically, when it comes to reducing the pulse rates of these participants, *Pet therapy* outperformed *Control* group ($p = .025$).

Incidentally, if all of the pairwise comparisons had produced *p* values that were greater than .05, then we would have accepted H_0 and rejected H_1.

Abstract

We recruited 90 participants who were diagnosed with acute stress disorder. We randomly assigned participants to one of three groups: Participants in Group 1 (the control group) received no supplemental treatment (for 30 minutes), those in Group 2 were provided a 30-minute pet therapy session with a certified therapy dog, and those in Group 3 engaged in a 30-minute guided meditation. After 30 minutes, the researcher recorded the pulse rate (beats per minute) of each participant.

The pulse rates were as follows: control group (M = 84.10, SD = 6.27), pet therapy group (M = 79.93, SD = 5.61), and meditation (M = 80.47, SD = 6.28).

The mean pulse rate of those in the pet therapy group was 4.17 lower than those in the control group (p = .025, α = .05*). This statistically significant finding suggests that pet therapy is a viable nonpharmaceutical supplemental treatment for those diagnosed with acute stress disorder. Considering that the meditation group had a pulse rate that was 3.63 lower than the control group (*p = .058, α = .05*), we intend to retain meditation in our next study.*

Occasionally, as in this example, a p value may be close to .05 (e.g., p = .058). In such instances, you may be tempted to comment that the .058 p level 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 somehow on its way to crossing the .05 finish line, but this is not at all the case. The .058 p value is actually a *static* variable, meaning that it is not in motion—the .058 p value is no more *approaching* .05 than it is *approaching* .06. Think of the .058 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 it (the .058 p value) 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: Consider the Number 4 parked right where it belongs on a number line. It is not *drifting* in any direction; it is not *approaching* 3 or 5.

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 5.9); minor variations in the normal distribution are acceptable. Occasionally, you may encounter data that are substantially skewed (Figure 5.10), bimodal (Figure 5.11), flat (Figure 5.12), or may have some other atypical distribution. In such instances, the **Kruskal-Wallis statistic test** is an appropriate alternative to the ANOVA test.

Figure 5.9 Normal.

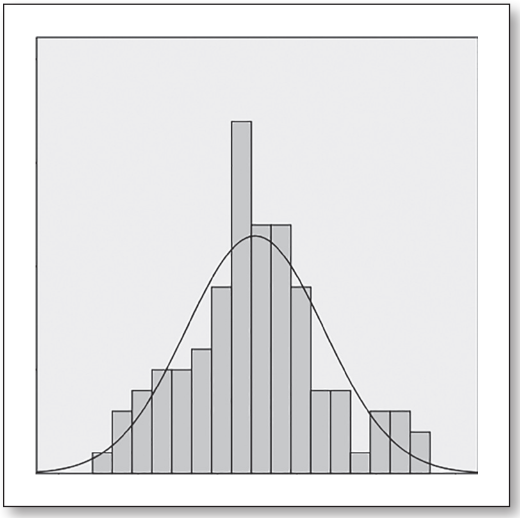


Figure 5.10 Skewed.

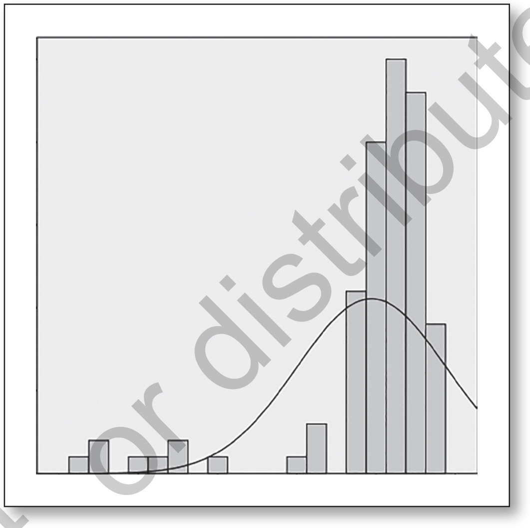


Figure 5.11 Bimodal.

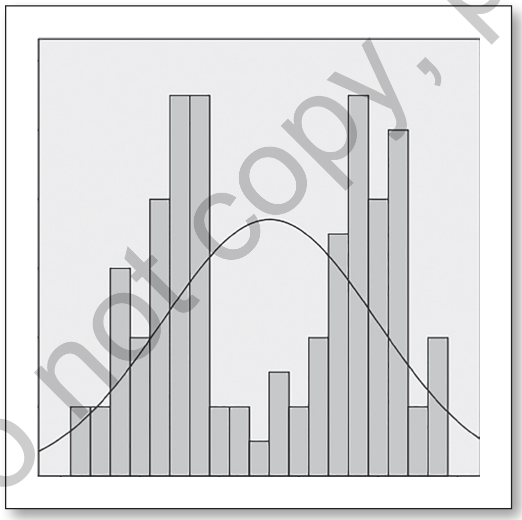
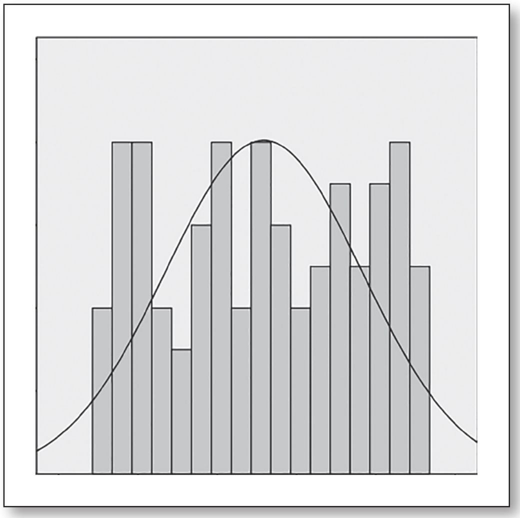


Figure 5.12 Flat.

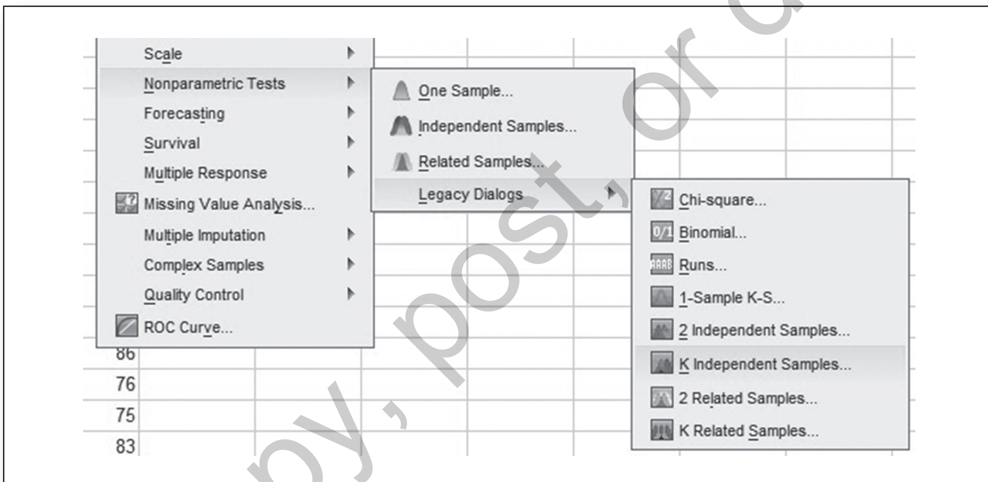


Test Run

For exemplary purposes, we will run the Kruskal-Wallis test using the same data set (**Ch 05 – 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 to the results produced by the Kruskal-Wallis test.

1. On the main screen, click on *Analyze, Nonparametric Tests, Legacy Dialogs, K Independent Samples* (Figure 5.13).

Figure 5.13 Ordering the Kruskal-Wallis test: Click on *Analyze*, *Nonparametric Tests*, *Legacy Dialogs*, *K Independent Samples*.



2. On the *Test for Several Independent Samples* menu, move *Pulse* to the *Test Variable List* window.
3. Move *Group* to the *Grouping Variable* box (Figure 5.14).
4. Click on *Group(? ?)*; then click on *Define Range*.
5. On the *Several Independent* Samples: Define Range submenu, for *Minimum, enter 1;* for *Maximum,* enter *3* (since the groups are numbered 1 [for *Control*] through 3 [for *Meditation*]) (Figure 5.15).
6. Click *Continue;* this will close this submenu.
7. On the *Tests for Several Independent Samples* menu, click on *OK*.

Figure 5.14 On the *Tests for Several Independent Samples* menu, move *Time* to *Test Variable List*, and move *Group* to the *Grouping Variable* box.

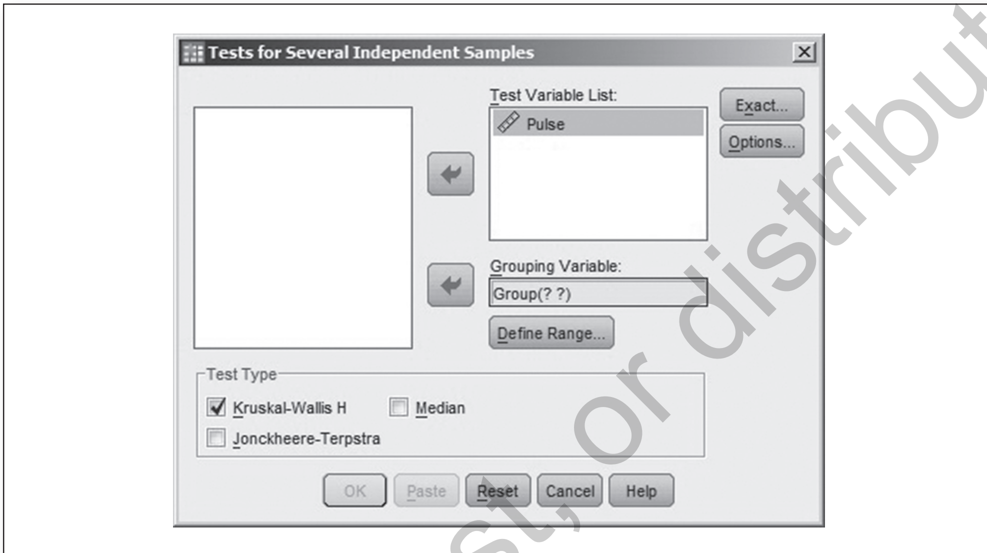
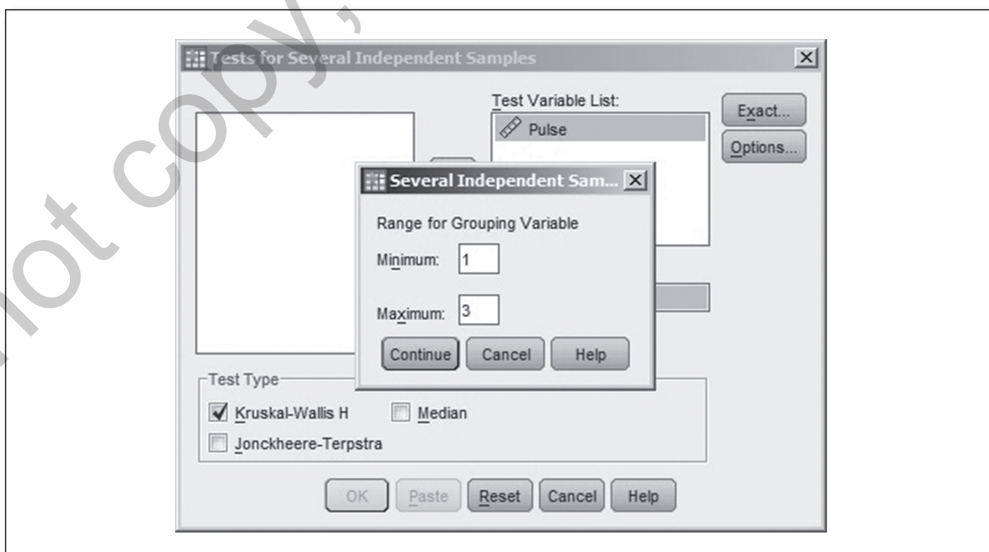


Figure 5.15 On the *Tests for Several Independent Samples* submenu, for *Minimum*, enter 1; for *Maximum*, enter 3.



Results

The Kruskal-Wallis result is found in the Test Statistics table (Table 5.9); the *Asymp. Sig.* statistic rendered a *p* value of .030; this is less than α (.05), so we would conclude that there is a statistically significant difference (somewhere) among the performances of the three *times*, but we still need to conduct pairwise (post hoc type) analyses to determine which group(s) outperformed which.

Table 5.9 Kruskal-Wallis *p* value = .030.

Test Statistics[a,b]

	Pulse
Chi-square	6.982
df	2
Asymp. Sig.	.030

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 *Control : Pet therapy,* then *Control : Meditation,* and finally *Pet therapy : Meditation*.

8. Click on the *Select Cases* icon.
9. On the *Select Cases* menu, click on ⦿ *If condition is satisfied* (Figure 5.16).
10. Click on *If*.
11. On the *Select Cases: If* menu, specify the pair of groups that you want selected (Figure 5.17):
    - *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 5.16 On the *Select Cases* menu, click on ⊙ *If condition is satisfied*, then click on *If*.

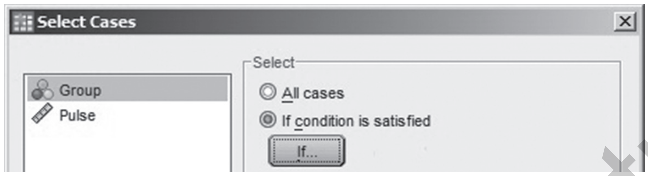


Figure 5.17 On the *Select Cases* menu, click on ⊙ *If condition is satisfied*, then click on *If*.

12. Click *OK*.
13. 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 5.10). NOTE: You can keep using the parameters specified from the previous run(s).

Table 5.10 Pairwise *p* values for the Kruskal-Wallis test (manually assembled).

Groups	*p*
Control : Pet therapy	.012*
Control : Meditation	.042*
Pet therapy : Meditation	.847

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

To finalize this discussion, consider Table 5.11, 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 5.11), remember that the ANOVA test produced an initial *p* value of .018 (which we read before the paired post hoc tests), whereas the Kruskal-Wallis produced an initial (overall) *p* value of .030. The differences in these *p* values are 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, then the Kruskal-Wallis test is considered a viable alternative.

Table 5.11 ANOVA and Kruskal-Wallis pairwise post hoc *p* values.

Groups	ANOVA *p*	Kruskal-Wallis *p*
Control : Pet therapy	.025*	.012*
Control : Meditation	.058	.042*
Pet therapy : Meditation	.938	.847

*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, pet therapy stood out as the best treatment with a mean pulse rate of 79.93; however, the meditation group came in as a close second with a mean pulse rate of 80.47. The .054 difference in mean pulse rates between these two groups is statistically insignificant ($p = .938$). This suggests that these two treatments produced very similar results when it comes to reducing the pulse rate of anxious participants. Despite the appeal of the therapy dogs, it may not be feasible or affordable to issue such pets to each participant. However, it would be fairly simple and affordable to provide each participant with a copy of the meditation recording so that they can practice meditation at home to facilitate longitudinal home care.

The point is that statistical analysis can provide precise results that can be used in making (more) informed decisions, yet 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 5.7 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 suitably (large) sample to facilitate normal distributions among all of the groups. Another consideration involves the documentation process. For example, 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/intermediatestats**).

NOTE: These practice exercises and data sets are the same as those in **Chapter 5: *t* Test and Mann-Whitney U test** except instead of the two-group designs, additional data has been included to facilitate ANOVA processing: Exercises 5.1 through 5.8 have three groups, and exercises 5.9 and 5.10 have four groups.

Exercise 5.1

You want to determine the optimal tutor-to-student ratio. Students seeking tutoring will be randomly assigned to one of three groups: Group 1 will involve each tutor working with only one student; in Group 2, each tutor will work with two students; and in Group 3, each tutor will work with five students. At the end of the term, students will be asked to complete the Tutor Satisfaction Survey, which renders a score from 0 to 100.

Data set: **Ch 05 – Exercise 01A.sav**

Codebook

Variable:	Group
Definition:	Group number (1 = One-to-one, 2 = Two-to-one, 3 = Five-to-one)
Type:	Categorical
Variable:	TSS
Definition:	Tutor Satisfaction Survey score (0 = Very unsatisfied . . . 100 = Very satisfied)
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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

d. Write an abstract under 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 05 – Exercise 01B.sav.**

Exercise 5.2

Clinicians at a nursing home facility want to see if giving residents a plant to tend to will help lower depression. To test this idea, the residents are randomly assigned to one of three groups: Those assigned to Group 1 will serve as the control group and will not be given a plant. Members of Group 2 will be given a small bamboo plant along with a card detailing care instructions. Members of Group 3 will be given a small cactus along with a card detailing care instructions. After 90 days, all participants will complete the Acme Depression Scale, which renders a score between 1 and 100 (1 = Low depression . . . 100 = High depression).

Data set: **Ch 05 – Exercise 02A.sav**

Codebook

Variable:	Group
Definition:	Group number
Type:	Categorical (1 = No plant, 2 = Bamboo, 3 = Cactus)

Variable: Depress

Definition: Acme Depression Scale

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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – Exercise 02B.sav.**

Exercise 5.3

A judge mandates that juvenile offenders who have priors be assigned to a trained delinquency prevention mentor. To assess this intervention, offenders will be randomly assigned to one of three groups: No mentor, a peer mentor who is 3 to 5 years older than the offender, or an adult mentor who is 10 or more years older than the offender. The following data will be gathered on each participant: Probation officer's compliance evaluation (0% . . . 100%).

Data set: **Ch 05 – Exercise 03A.sav**

Codebook

Variable: Group

Definition: Mentor group assignment

Type: Categorical (1 = No mentor, 2 = Peer mentor, 3 = Adult mentor)

Variable: Probation_compliance

Definition: Probation officer's overall assessment of the youth's probation compliance

Type: Continuous (0 = Completely non-compliant . . . 100 = Completely compliant)

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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – Exercise 03B.sav.**

Exercise 5.4

In an effort to determine the effectiveness of light therapy to alleviate depression, you recruit a group of individuals who have been diagnosed with depression. The participants 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 05 – Exercise 04A.sav**

Codebook

Variable:	Group
Definition:	Group number
Type:	Categorical (1 = No light therapy, 2 = Light therapy: even days, 3 = Light therapy: every day)
Variable:	Mood
Definition:	Acme Mood Scale
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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – Exercise 04B.sav**.

Exercise 5.5

To assess the workplace benefits of providing paid time off (PTO), the Human Resources (HR) Department implements and evaluates different PTO plans at each of the company's three sites: Site 1 will serve as the control group; employees at this site will continue to receive 2 weeks of PTO per year. Employees at Site 2 will receive 2 weeks of PTO per year plus the fourth Friday of each month off (with pay). Employees at Site 3 will receive 3 weeks of PTO per year. The HR Department will use a web-based survey to gather the following data from all employees: score on the Acme Morale Scale (1 = extremely low morale . . . 25 = extremely high morale).

Data set: **Ch 05 – Exercise 05A.sav**

Codebook

Variable:	Site
Definition:	Work site
Type:	Categorical (1 = 2 Weeks PTO, 2 = 2 Weeks PTO + fourth Fridays off, 3 Weeks PTO)
Variable:	Morale
Definition:	Score on Acme Morale Scale
Type:	Continuous (1 = Extremely low morale . . . 25 = Extremely high morale)

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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – Exercise 5B.sav.**

Exercise 5.6

It is thought that exercising early in the morning will provide better energy throughout the day. To test this idea, participants 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 05 – Exercise 06A.sav**

Codebook

Variable:	Group
Definition:	Walking group assignment
Type:	Categorical (1 = No walking, 2 = Walking: 30 Minutes, 3 = Walking: 60 minutes)

Variable:	Energy
Definition:	Acme End-of-the-Day Energy Scale
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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – exercise 06B.sav**.

Exercise 5.7

A political consulting firm wants to determine the characteristics of voters when it comes to issues involving alternative energy. The researchers recruit a group of participants and randomly assign them to one of three groups: Group 1 will be the control group, and they will not be exposed to any advertising materials; Group 2 will be shown a print advertisement that will be used in a postal mailing; and Group 3 will be shown a video advertisement that will be aired on television. Finally, each participant will indicate his or her voting intentions for Proposition 86, which involves tax deductions for hybrid cars on a 1 to 7 scale (1 = Will definitely vote no . . . 7 = will definitely vote yes).

Data set: **Ch 05 – Exercise 07A.sav**

Codebook

Variable:	Group
Definition:	Advertising media
Type:	Categorical (1 = Control, 2 = Print, 3 = Video)

Variable:	Prop_86
Definition:	Likely voting decision on tax deductions for hybrid cars
Type:	Continuous (1 = Will definitely vote no . . . 7 = Will definitely vote yes)

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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – exercise 07B.sav**.

Exercise 5.8

A team of educational researchers wants to assess traditional classroom instruction compared to online options. Students who are enrolled in a course will be randomly assigned to one of three sections: Students in Section 1 will take the class in a traditional classroom. Students in Section 2 will take the course online with an interactive video cast of the instructor wherein students can ask the instructor questions during the session. Students in Section 3 will take the course online and view a prerecorded video of the professor delivering the lecture. The researchers will gather the course grade of each student (0% . . . 100%).

Data set: **Ch 05 – Exercise 08A.sav**

Codebook

Variable:	Section
Definition:	Learning modality
Type:	Categorical (1 = Classroom, 2 = Online live interactive, 3 = Online prerecorded video)

Variable: Grade

Definition: Final grade in course

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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – Exercise 08B.sav**.

NOTE: Exercises 9 and 10 involve four groups each.

Exercise 5.9

The Acme Company claims that its new reading lamp increases reading speed, and 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 05 – 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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – Exercise 09B.sav.**

Exercise 5.10

Due to numerous complications involving missed medication dosages, you implement a study to determine the best strategy for enhancing medication adherence. Patients who are on a daily medication regime will be recruited, will receive a complimentary 1-month dosage of their regular medication(s), and will be randomly assigned to one of four groups. Group 1 will serve as the control group (no treatment); Group 2 will participate in a 1-hour in-person pharmacist-administered medication adherence workshop; Group 3 will receive a text message reminder with a picture of the drug (e.g., *It's time to take one tablet of Drug A*); Group 4 will attend the medication adherence workshop and also receive text messages. At the end of 1 month, participants will present their prescription bottle(s); you will count the remaining pills and calculate the dosage adherence percentage (e.g., 0 pills remaining = 100% adherence).

Data set: **Ch 05 – Exercise 10A.sav**

Codebook

Variable:	Group
Definition:	Group number
Type:	Categorical (1 = Control, 2 = Rx workshop, 3 = Texts, 4 = Rx workshop and texts)

Variable:	RxAdhere
Definition:	Percentage of medication adherence
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 (*n*s, means, and Sig. [*p* value], hypotheses resolution).

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

Repeat this exercise using **Ch 05 – Exercise 10B.sav**.