Introductory Statistics Using SPSS

Author: Herschel Knapp

Pub. Date: 2022

Product: Sage Research Methods

DOI: https://doi.org/10.4135/9781071878910

Methods: T-test, Descriptive statistics, Normal distribution

Disciplines: Psychology, Mathematics

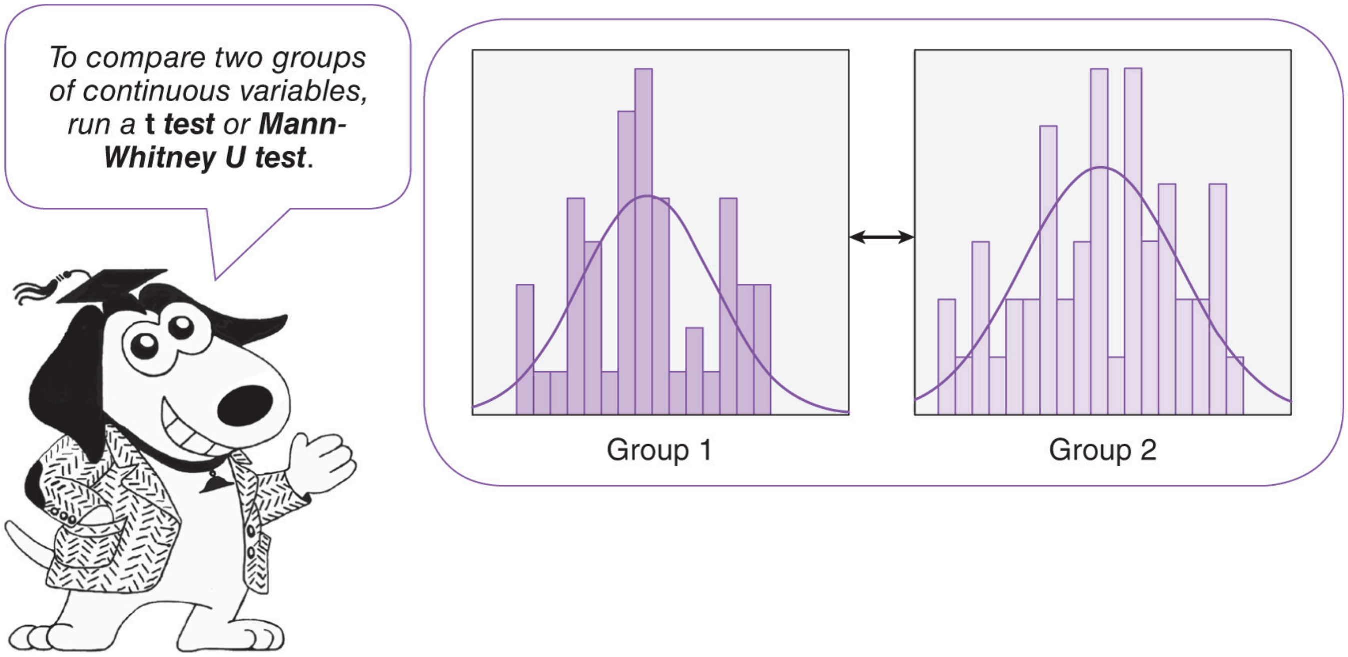
Access Date: July 19, 2024

Publisher: SAGE Publications, Inc

City: Thousand Oaks

Online ISBN: 9781071878910

t Test and Mann-Whitney *U* Test



The difference between a violin and a viola is that a viola burns longer.

—Victor Borge

Learning Objectives

Upon completing this chapter, you will be able to:

- Determine when it is appropriate to run a *t* test
- Verify that the data meet the criteria for *t* test processing: normality, *n*, and homogeneity of variance
- Order a *t* test
- Interpret the test results
- Comprehend the α and *p* values
- Resolve the hypotheses
- Know when and how to run and interpret the Mann-Whitney *U* test
- Document the results in plain English
- Understand the implications of Type I and Type II errors

- Apply techniques for reducing the likelihood of committing Type I and Type II errors

NOTE: From here forward, the μ character will be used to symbolize the mean.

Videos

The videos for this chapter are **Ch 05 - t Test.mp4** and **Ch 05 - Mann-Whitney U Test.mp4**. These videos provide overviews of these tests, instructions for carrying out the pretest checklist, running the tests, and interpreting their results using the data set **Ch 05 - Example 01 - t Test and Mann-Whitney U Test.sav**.

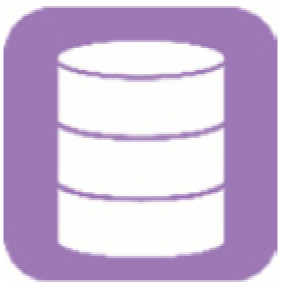
Overview—*t* Test

The *t* test is one of the most common and versatile statistical tests in the realm of experimental research and survey methodology. The *t* test is used when there are two groups, wherein each group renders a continuous variable for the outcome (e.g., height, age, weight, number of teeth, bank account balance, IQ score, score on a depression assessment instrument, pulse rate, test score, number of crimes, typing speed).

In the most basic experimental settings, the design consists of two groups: a control group, which gets nothing, a placebo, or treatment as usual, and a treatment group, which gets the innovative intervention that is the focus of the study.

We can compute the mean for each group, and we would not expect the two means to be identical; they would likely be different. The *t* test answers the question “Is there a statistically significant difference between μ(control) and μ(treatment)?” In other words, the result of the *t* test helps us determine if one group *substantially* outperformed the other, or if the differences between the means are essentially *incidental*.

In cases where the three pretest criteria are not satisfied for the *t* test, the Mann-Whitney *U* test, which is conceptually similar to the *t* test, 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 or (2) Text with illustrations.

Research Question

Do illustrations in the instructions facilitate quicker assembly of the Able Chair?

Groups

A researcher recruits a total of 70 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 two groups on an alternating basis: 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.

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) fails, indicating that on average, participants who are issued the assembly instructions with illustrations 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 for these two groups. The alternative hypothesis (H_1) states that, on average, those who use the instructions with illustrations assemble the chair faster than those who use the text-only instructions:

> H_0: There is no difference in assembly time between the groups.
> H_1: The Text with illustrations group assembles the chair faster than the Text group.

Data Set

Use the following data set: **Ch 05 - Example 01 - t Test and Mann-Whitney U test.sav.**

Codebook

> Variable: group
> Definition: Group assignment
> Type: Categorical (1 = Text, 2 = Text with illustrations)

Variable: time

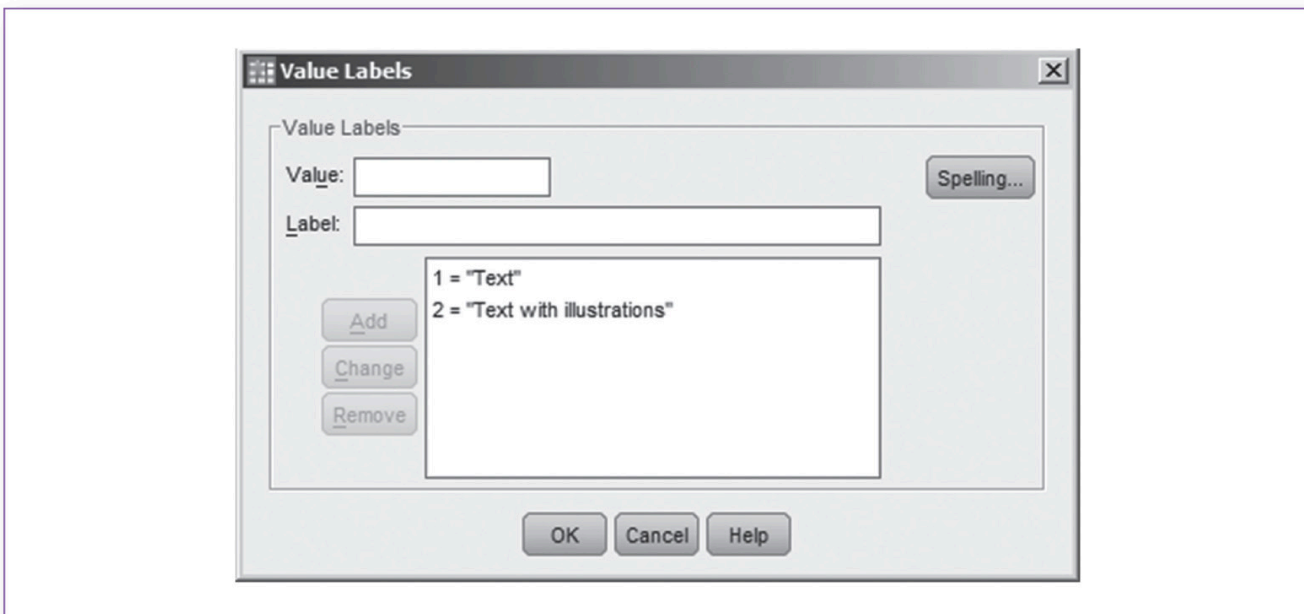
Definition: Minutes required to assemble the Able Chair

Type: Continuous

NOTE: In this data set, records (rows) 1 through 35 are for Group 1 (Text), and records 36 through 70 are for Group 2 (Text with illustrations). 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 corresponding categorical labels have been assigned: 1 for Text and 2 for Text with illustrations (Figure 5.1).

Figure 5.1 Value labels for a t test analysis.



Pretest Checklist

t Test Pretest Checklist

- ☑ 1. Normality[a]
- ☑ 2. *n* quota[b]
- ☑ 3. Homogeneity of variance[b]

a. Run prior to *t* test.

b. Results produced upon *t* test run.

The statistical pretest checklist is akin to looking both ways before you cross the street; certainly you could cross the street without looking, but you would probably wind up in much better shape if you looked first. In terms of statistical tests, certainly you could run the statistical test without tending to the pretest checklist, but you may unknowingly generate misleading findings.

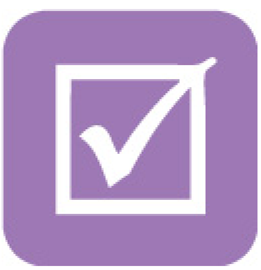
The formulas that compose each statistical test require that the source data meet a unique set of criteria in order for that test to operate properly. These criteria are referred to as *assumptions*—we *assume* that the data meet the criteria specified by the test at hand. Actually, we need to do more than just passively *assume* that the data are suitable for processing; we need to *actively assess* the source data before proceeding with the test.

When the tests on the pretest checklist (statistical assumptions) are satisfied, we can consider the statistical results to be relatively robust. If there are minor deviations in these criteria, we could still proceed with the analysis, but we would be a bit less confident in the solidity of our findings. In the interest of proper scientific ethics and the principles of full disclosure, it would be appropriate to mention any such (statistical) shortcomings when discussing the results.

In instances where one or more of the specified pretest criteria are substantially not satisfied, the better option is to use the nonparametric version of the test, in this case, the Mann-Whitney *U* test, which was mentioned earlier. This notion pertains to the unique pretest checklists associated with the other tests covered in this text as well.

The pretest criteria for running a *t* test involve checking the data for **(a) normality**, **(b) *n* quota**, and **(c) homogeneity** (pronounced *hoe-moe-juh-nay-it-tee*) **of variance**.

Pretest Checklist Criterion 1—Normality



Checking for normality involves producing a histogram with a normal curve for each of the two groups. In this instance, you would click on the *Select Cases* icon to select the records pertaining to the Text group; the selection criteria would be *group* = 1. Next, run a histogram (with normal curve) on the variable *time*. Then repeat the process for the Text with illustrations group (*group* = 2). For more details on this procedure, please 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.

This will produce two histograms with normal curves—one for *time* in the Text group and the other for *time* in the Text with illustrations group. The histograms should resemble the graphs shown in Figures 5.2 and 5.3.

As we read these two histograms, set aside the X and Y scales and the possible irregularities among the bars; our attention is focused primarily on the shape of the normal curve that is superimposed over the bars. We are looking for normality (symmetry) in these two curves. Although the normal curve in Figure 5.2 is shorter and fatter than the normal curve in Figure 5.3, in terms of normality, this is not an issue. The critical thing to observe is that both normal curves are sufficiently symmetrical—in other words, if you sliced each curve vertically down the middle, the left side would resemble a mirror image of the right side. Sometimes this normal curve is aptly referred to by its characteristic shape as a “bell curve.” In this example, we see that both curves are symmetrical; there is no notable skew in either curve. Hence, we would say that the criteria of normality are satisfied for both Text and Text with illustrations.

Figure 5.2 Histogram of time for Group 1: Text.

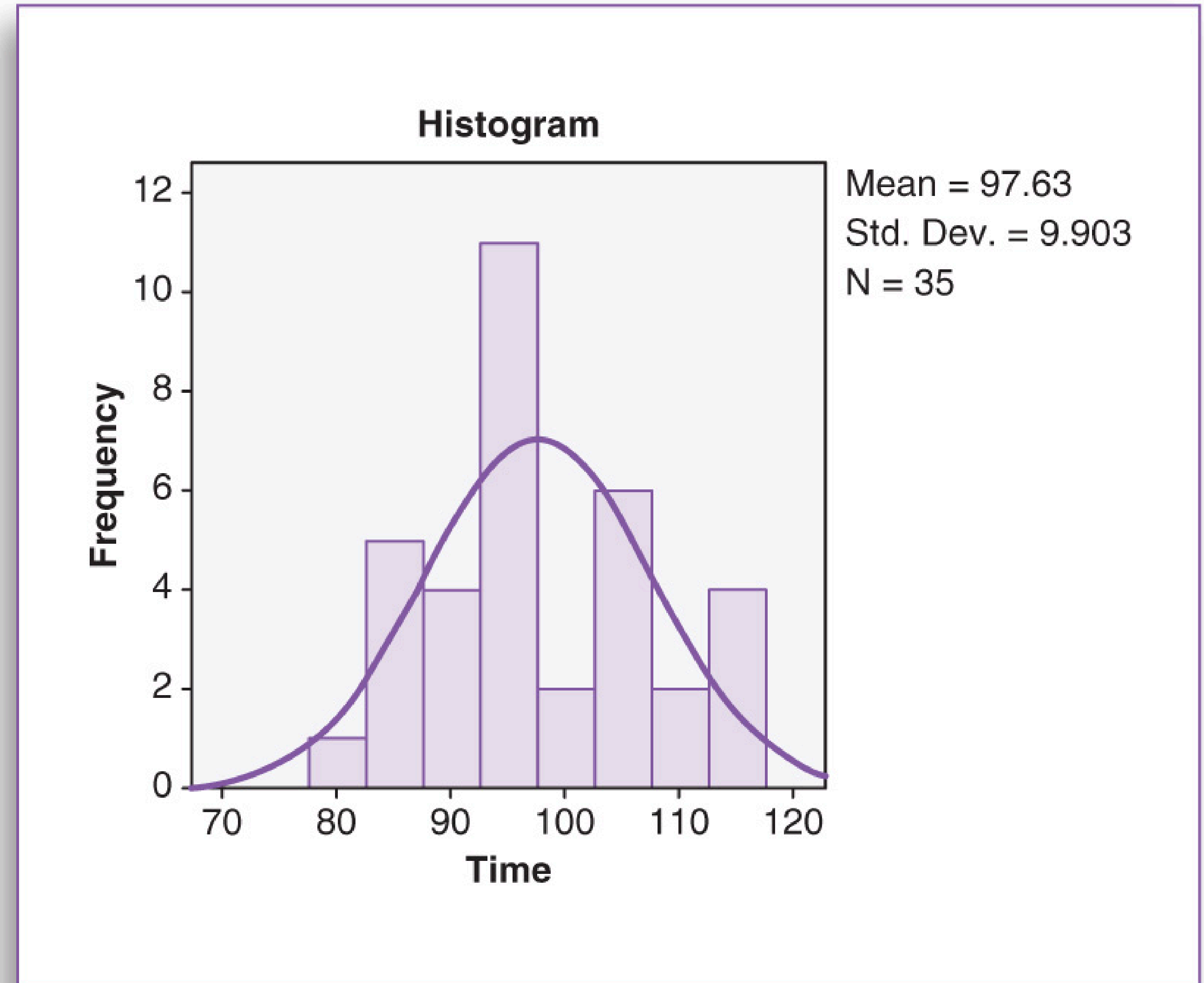
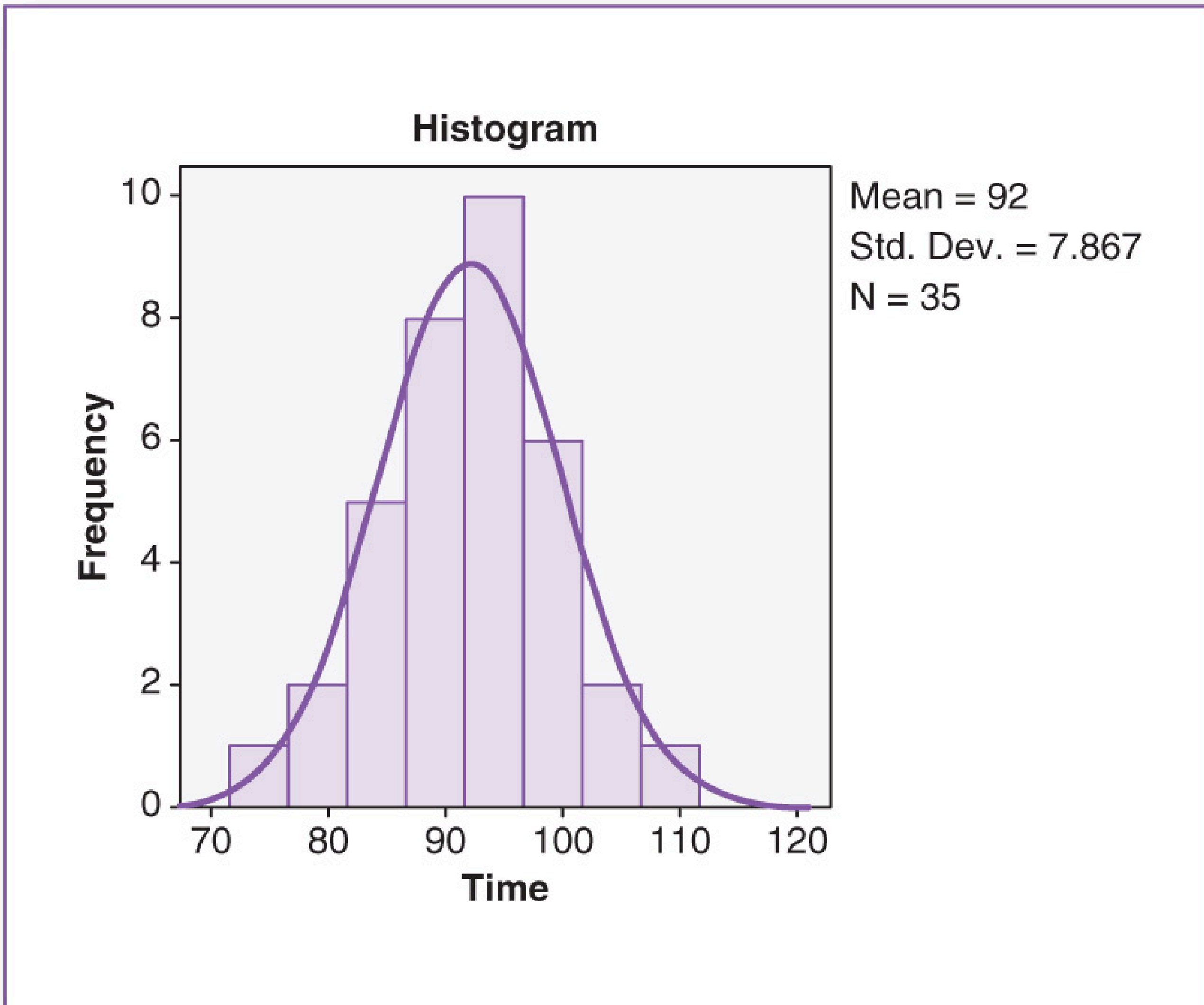


Figure 5.3 Histogram of time for Group 2: Text with illustrations.



The remaining two pretest criteria, *n* quota and homogeneity of variance, will be processed during the **Test Run** section and finalized in the **Results** section.

Pretest Checklist Criterion 2—n Quota

Technically, you can process a *t* test with an *n* of any size in each group, but when the *n* is at least 30 in each group, the result of the *t* test is considered more robust. We will see the *n* reported for each group in the **Results** section, when we examine the findings produced by the *t* test.

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) of 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 5.2: $9.0903^2 = 82.6336$), and the variance for *time* in the Text with illustrations group is 61.89 (derived from Figure 5.3: $7.867^2 = 61.8897$). Clearly, 82.63 is not more than twice 61.89, so we would expect that the homogeneity of variance test would pass.

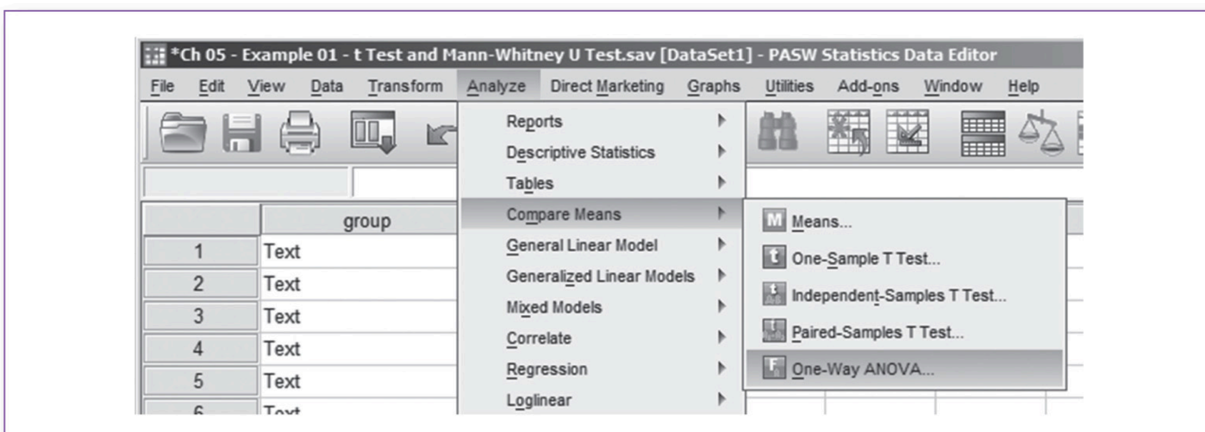
In SPSS, the homogeneity of variance test is an option selected during the actual run of the *t* test. If the homogeneity of variance test renders a significance (*p*) value that is greater than .05, this suggests that there is no statistically significant difference between the variance of one group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

Test Run

You may have noticed a variety of *t* test options on the *Analyze*, *Compare Means* pull-down menu. We will be using the *One-Way ANOVA* window to process the *t* tests. The reasoning for this option is threefold: (a) This is an easier window to fill out, (b) it will produce the desired *t* test results, and (c) this will prepare you to efficiently run the ANOVA (analysis of variance) in the next chapter. Note that the *t* test is basically a two-group ANOVA.

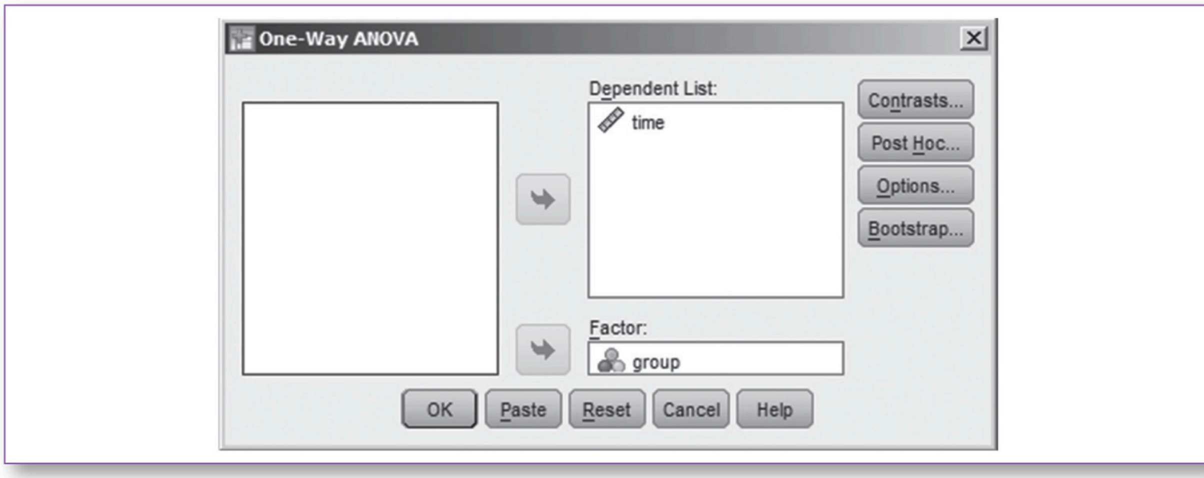
1. On the main screen, click on *Analyze*, *Compare Means*, *One-Way ANOVA* (Figure 5.4).

   Figure 5.4 Running a t test.



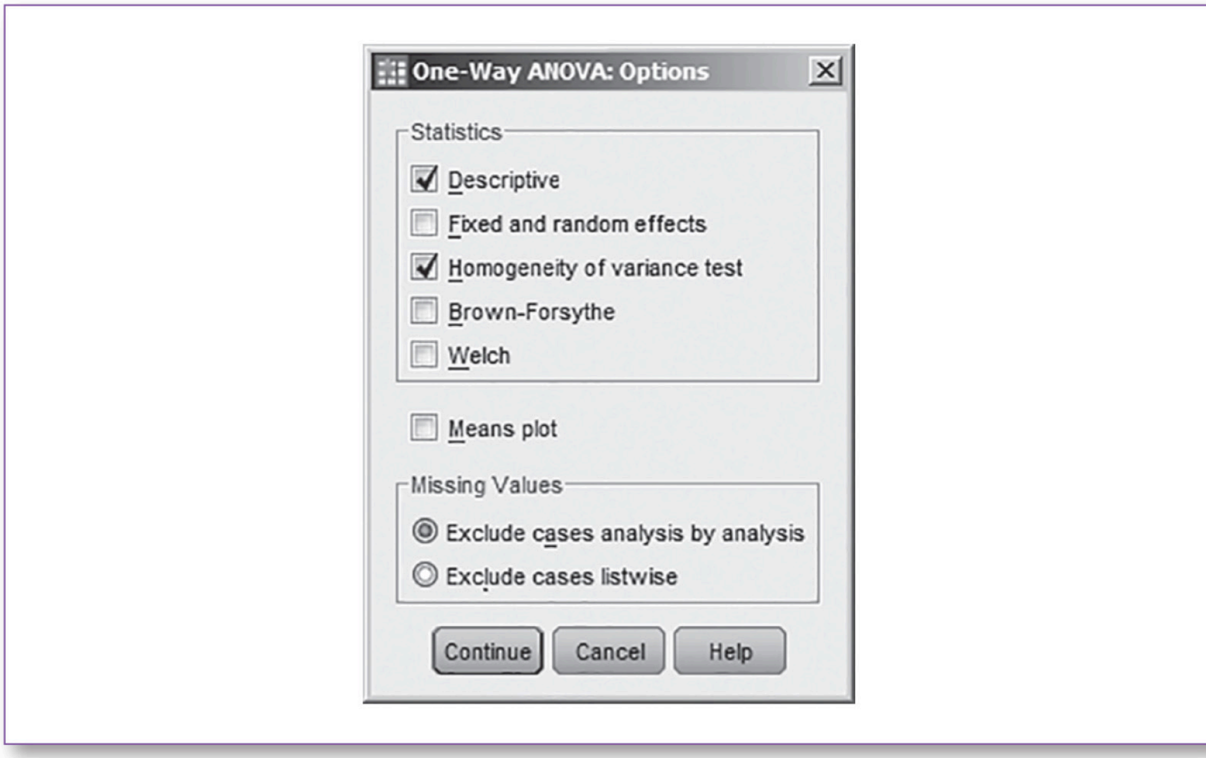
2. On the *One-Way ANOVA* window, move the continuous variable that you wish to analyze (*time*) into the *Dependent List* panel, and move the variable that contains the categorical variable that specifies the groups (*group*) into the *Factor* panel (Figure 5.5).

Figure 5.5 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 5.6). This will take you back to the *One-Way ANOVA* window.

Figure 5.6 The One-Way ANOVA: Options window.



4. In the *One-Way ANOVA* window (Figure 5.5), click on the *OK* button, and the test will process.

Results

Pretest Checklist Criterion 2—n Quota

We will begin this section by completing the remaining pretest checklist criteria; we still need answers regarding the *n* and the homogeneity of variance.

The *Descriptives* table (Table 5.1) shows that each group has an *n* of 35. This satisfies the *n* quota, indicating that the ANOVA test becomes more robust when the *n* for each group is at least 30.

Table 5.1 Descriptive Statistics Results.

Table 5.1 Descriptive Statistics Results.

Descriptives

time

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Text	35	97.63	9.903	1.674	94.23	101.03	80	115
Text with illustrations	35	92.00	7.867	1.330	89.30	94.70	74	110
Total	70	94.81	9.319	1.114	92.59	97.04	74	115

Pretest Checklist Criterion 3—Homogeneity of Variance

The last column of the *Test of Homogeneity of Variances* table (Table 5.2) shows a significance (Sig.) of

.051—this is the *p* value. Since this is greater than the α level of .05, this tells us that there is no statistically significant difference between the variances in the *time* variable for the Text group compared with the Text with illustrations group. We would conclude that the criterion of homogeneity of variance has been satisfied. (The significance and α level are clarified in the following section.)

Table 5.2 Homogeneity of Variance Test Results.

Table 5.2 Homogeneity of Variance Test Results.

Test of Homogeneity of Variances

time

Levene Statistic	df1	df2	Sig.
3.933	1	68	.051

p Value

At this point, you have probably noticed column 3 in Table 5.1, which shows that the mean *time* for the Text with illustrations group (92.00) is lower than that for the Text group (97.63). On the basis of these means, you may hastily conclude that Text with illustrations is the best, since the participants in that group assembled the chair more quickly than those in the text group, but in statistics, the world is not so simple.

Statisticians recognize and actively acknowledge that we do not live in a perfect world; no matter how hard we try to conduct quality investigations, the scientific process can be a messy proposition littered with multiple confounding variables—conditions for which we cannot fully control or account for that can influence the outcome variable(s).

In our simple example, judging by the mean times of the two groups, it looks like the Text with illustrations group outperformed the Text group, and this may in fact turn out to be the case, but other factors may have contributed to the differences observed between the (mean) times of these two groups. For example, maybe the random distribution process unexpectedly routed more proficient people to the Text with illustrations group; maybe the majority of unmotivated people were unintentionally routed to the Text group; maybe the members of the Text group had poorer dexterity than those in the Text with illustrations group. Any number of these or other factors that we may not know about may have been occurring over the course of this study.

As you can see, there is virtually no limit to the variety of confounding variables that could potentially influence the outcome of a study. Since we cannot fully account for or compensate for such confounds, we know we do not have a perfect experimental setting. Hence, we do not speak of our statistical findings with *absolute certainty*; rather, we speak of how much *confidence* we have in our findings.

The key question in this case is, *How certain can we be that the 5.63-minute difference we detected between the group means (μ[Text with illustrations] = 92.00 and μ[Text] = 97.63) is actually due to the genuine superiority of Text with illustrations and not due to chance alone?* In other words, we want a metric that will tell us how likely it is that we would detect this result (the 5.63-minute difference in the means) if Text with illustrations was actually no different from the Text-only method. This number is known as the *significance level*, represented by the letter *p*.

Here is how the significance (*p*) value works: Look at the last column of Table 5.3; the significance score is .010; this is the *p* value. This tells us that we would expect to see the 5.63-minute difference in the group times about 1% of the time if it were occurring by (random) chance alone. In other words, based on the data gathered, if Text with illustrations is exactly as effective as Text, we would see the Text with illustrations group outperform the Text group by 5.63 minutes about 1% of the time.

Table 5.3 t Test Results Comparing Times of Text and Text With Illustrations.

Table 5.3 *t* Test Results Comparing Times of Text and Text With Illustrations.

ANOVA

time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	554.414	1	554.414	6.933	.010
Within Groups	5438.171	68	79.973		
Total	5992.586	69			

Since the *p* value tells us how often we would be accepting the intervention as effective, when in fact it really is not, the lower the *p* value, the more significant the findings.

To exemplify the point further, suppose this experiment produced a *p* value of .01. This tells us that if Text with illustrations were in fact equivalent to Text, and we ran this experiment 100 times, 1 of those iterations would, through random chance, produce results wherein Text with illustrations would outperform Text. If the *p* value was .001, this indicates that we would have to run this experiment 1,000 times to see an instance where Text with illustrations outperforms Text (merely due to random chance). Essentially, the lower the *p* value, the less likely it is that the findings (differences between the means of the groups) are occurring merely due to random chance, suggesting the stronger likelihood that the observed differences are due to the intervention—in this case, the effectiveness of Text with illustrations.

In the next section, we will see how we use the *p* value to determine which hypothesis to reject and which to accept.

Hypothesis Resolution

We need to have a way of using the *p* value to guide us in making decisions about our pending hypotheses:

> H_0: There is no difference in assembly time between the groups.
>
> H_1: The Text with illustrations group assembles the chair faster than the Text group.

α Level

To do this, before we embark on our research process, we draw a somewhat arbitrary numerical line in the sand, known as the alpha (α) level. Typically, the α level is set to .05. Think of the α level as a sort of statistical significance threshold—any *p* (significance) value that is .05 or less is considered statistically significant—hence, we reject H_0, which states that there is no significant difference between the groups. If the *p* value is greater than .05, then the differences between the means are not considered statistically significant—hence, we do not reject H_0. This will guide us in making our decisions regarding the hypotheses.

p Value Summary

- If $p \leq \alpha$, there is a statistically significant difference; reject H_0.
- If $p > \alpha$, there is no statistically significant difference; do not reject H_0.

NOTE: $\alpha = .05$ (.05 is the typical value used; some more stringent studies may use .01 or lower).

Since the *p* value is .010, which is less than or equal to α (.05), we would determine that there is a statistically significant difference between the times—specifically, that 92.00 minutes is statistically significantly less time than 97.63 minutes. To finalize the hypotheses, we would reject H_0 and accept H_1:

> **REJECT H_0:** There is no difference in assembly time between the groups.
>
> **ACCEPT H_1:** The Text with illustrations group assembles the chair faster than the Text group.

Documenting Results

Although it is essential to comprehend the meaning of the key values in the statistical reports, it would be inappropriate to simply present the figures in a results section without providing a concise narrative. While all figures below are technically correct, try to avoid documenting your findings as such:

Accurate but Inappropriate Numerical Statistical Abstract

> Text: *n* = 35, μ = 97.63 (SD = 9.90)
>
> Text with illustrations: *n* = 35, μ = 92.00 (SD = 7.87)
>
> *p* = .010, α = .05, therefore, use Text with illustrations

While the above data may be useful in assembling a table, it is important that you become proficient at translating your methodology and numerical findings into a brief textual abstract detailing the story of the study specifying the research question, along with an overview of how you got from the research question to the results.

Appropriate Textual Statistical Abstract

We recruited 70 participants on an individual basis to assemble the Able Chair, with the necessary tool (a screwdriver). Half of the participants were provided assembly instructions consisting of text only, and the other half were given instructions containing text with illustrations.

Those who used the text-only instructions took an average of 97.63 (SD = 9.90) minutes to assemble the chair, whereas those who used the instructions with illustrations assembled the chair in an average of 92.0 (SD = 7.87) minutes.

A t *test revealed that this 5.63-minute difference is statistically significant* (p = *.010,* α = *.05), suggesting that the instructions containing the text with illustrations facilitate quicker assembly of the Able Chair compared with the text-only instructions.*

In addition to the full manuscript, scientific journals also require authors to submit an abstract that tells the overall story of the study and key findings; usually the limit for the abstract is about 200 words. While initially it can be a challenge to write technical information so concisely, this is a worthy skill to develop. The above abstract is about 120 words.

NOTE: In the example processed in this chapter, we saw that the *t* test assessed the means of the two groups and revealed that the mean time for the treatment group (Text with illustrations) was statistically significantly lower than the mean time for the control group (Text), signifying the success of the illustrations in shortening the assembly time. As you will see in the exercises for this chapter, the *t* test is equally effective in detecting statistically significant differences when the mean of the treatment group is higher than the mean of the control group. For example, instead of the treatment (Text with illustrations) that is designed to *decrease* assembly time, in Exercise 2 in the Practice Exercises, the treatment (mentorship) is designed to *increase* academic grades of the mentee. Remember: The *t* test is designed to simply detect statistically significant differences between the means of two groups; it does not matter which group mean is higher and which is lower.

Type I And Type II Errors

The world is imperfect; despite all best efforts, errors can occur in virtually any realm no matter how careful you are. Consider the two types of errors that can occur in a legal verdict:

Error 1: The court finds the defendant guilty when, in fact, he or she is actually not guilty.
Error 2: The court finds the defendant not guilty when, in fact, he or she actually is guilty.

These same two types of errors can happen in statistics. Consider this standard set of hypotheses:

H_0: There is no significant difference between the groups ($p > .05$; the treatment failed).
H_1: There is a statistically significant between the groups ($p \leq .05$; the treatment worked).

Type I Error

A Type I error (also known as alpha [α] error) occurs when the findings indicate that there is a statistically significant difference between two variables (or groups) ($p \leq .05$) when, in fact, on the whole, there actually is not, meaning that you would erroneously reject the null hypothesis. The consequence is that you would conclude that treatment was effective when, in fact, on the whole, it was not. This is connected with the *p* value; a *p* value of .05 means that there is a 5% chance that you have committed a Type I error—hence, the lower the *p* value, the less likely it is that you have committed a Type I error. A Type I error can be thought of as the court finding the defendant guilty when, in fact, he or she is actually not guilty.

Type II Error

A Type II error (also known as beta [β] error) occurs when the findings indicate that there is no statistically significant difference between two variables (or groups) ($p > .05$) when, in fact, on the whole, there actually is, meaning that you would erroneously accept the null hypothesis. The consequence is that you would conclude that the treatment was ineffective when, in fact, on the whole, it was. Sample size is inversely related to Type II errors—the larger the sample size, the lower the likelihood of committing a Type II error. A Type II error can be thought of as a court finding the defendant not guilty when, in fact, he or she actually is guilty.

There is no formal metric you can run that will tell you if you have a Type I or Type II error on hand; they are just characteristics endemic in the realm of statistical testing. The point to keep in mind is that even if a statistical test produces a statistically significant *p* value (e.g., $p \leq .05$), this does not mean that you have solid evidentiary proof of anything; at best, you have reduced uncertainty. Essentially, a *p* value of .05 means that if the effect of the treatment group were the same as that of the control group (null intervention), we would see this (anomalous) statistical outcome, where the treatment group outperforms the control group, just by chance, about 5% of the time. Since *p* never reaches zero, there is always some level of uncertainty in statistical findings.

Occasionally, SPSS will produce results wherein the Sig. (p) value is .000. In such instances, the p value is so low that rounding it to three decimal digits produces the ".000" readout. In documenting such an occurrence, instead of writing "p = .000" or "p = 0," it is customarily documented as "p < .001."

Remember: Statistics is not about *proving* or *disproving* anything. Statistics is about reducing uncertainty—there is always some margin of error, no matter how small. The notion of Type I and Type II errors pertains to all other tests covered in the chapters that follow.

Overview—Mann-Whitney *U* Test

The Mann-Whitney U test is best thought of as a variation on the t test. As mentioned earlier, one of the pretest criteria that must be met prior to running a t test states that the data from each group must be normally distributed (Figure 5.7); minor variations in the normal distribution are acceptable. Occasionally, you may encounter data that are substantially skewed (Figure 5.8), bimodal (Figure 5.9), or flat (Figure 5.10) or that have some other atypical distribution. In such instances, the Mann-Whitney U statistic is an appropriate alternative to the t test.

Figure 5.7 Normal distribution.

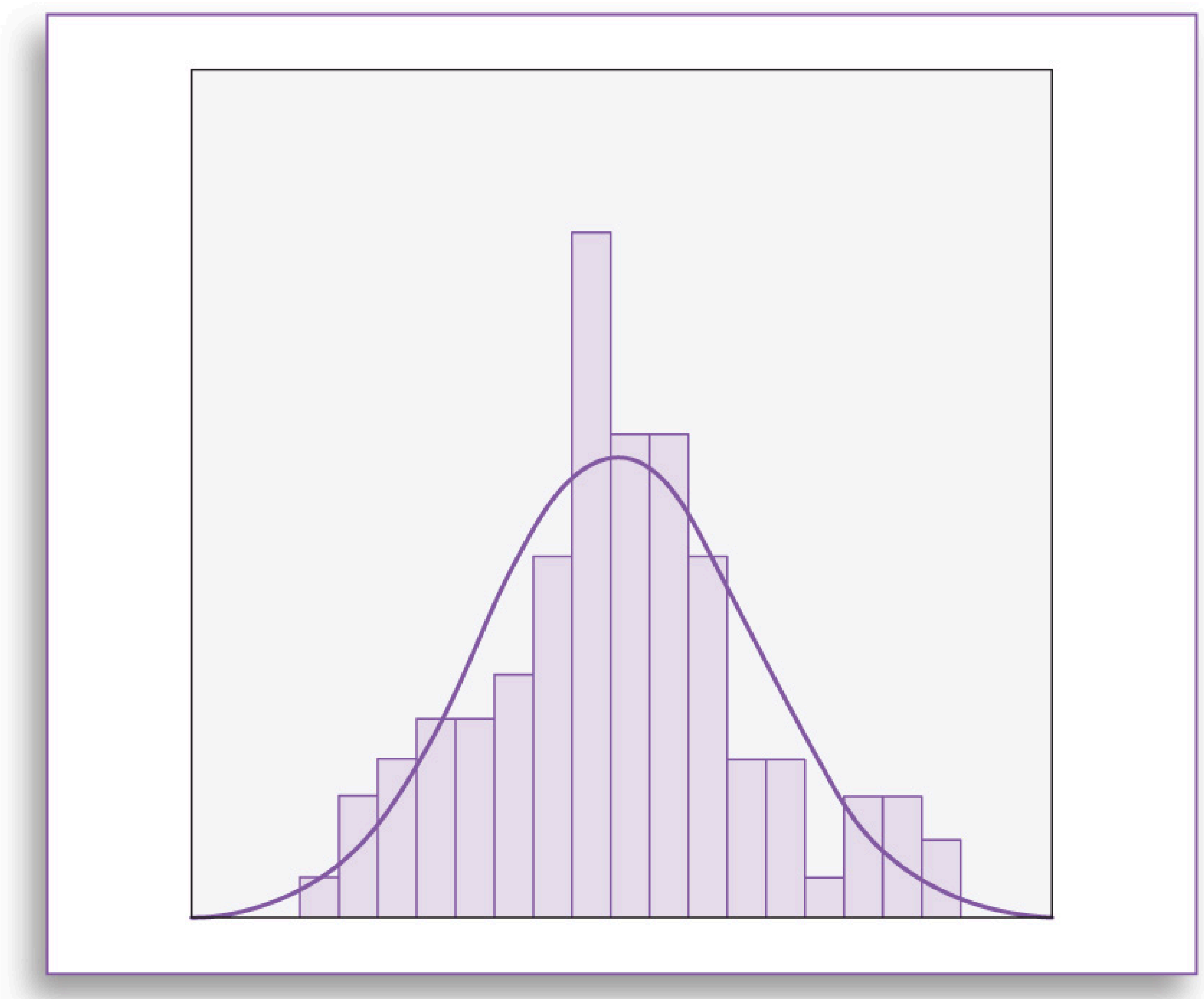


Figure 5.8 Skewed distribution.

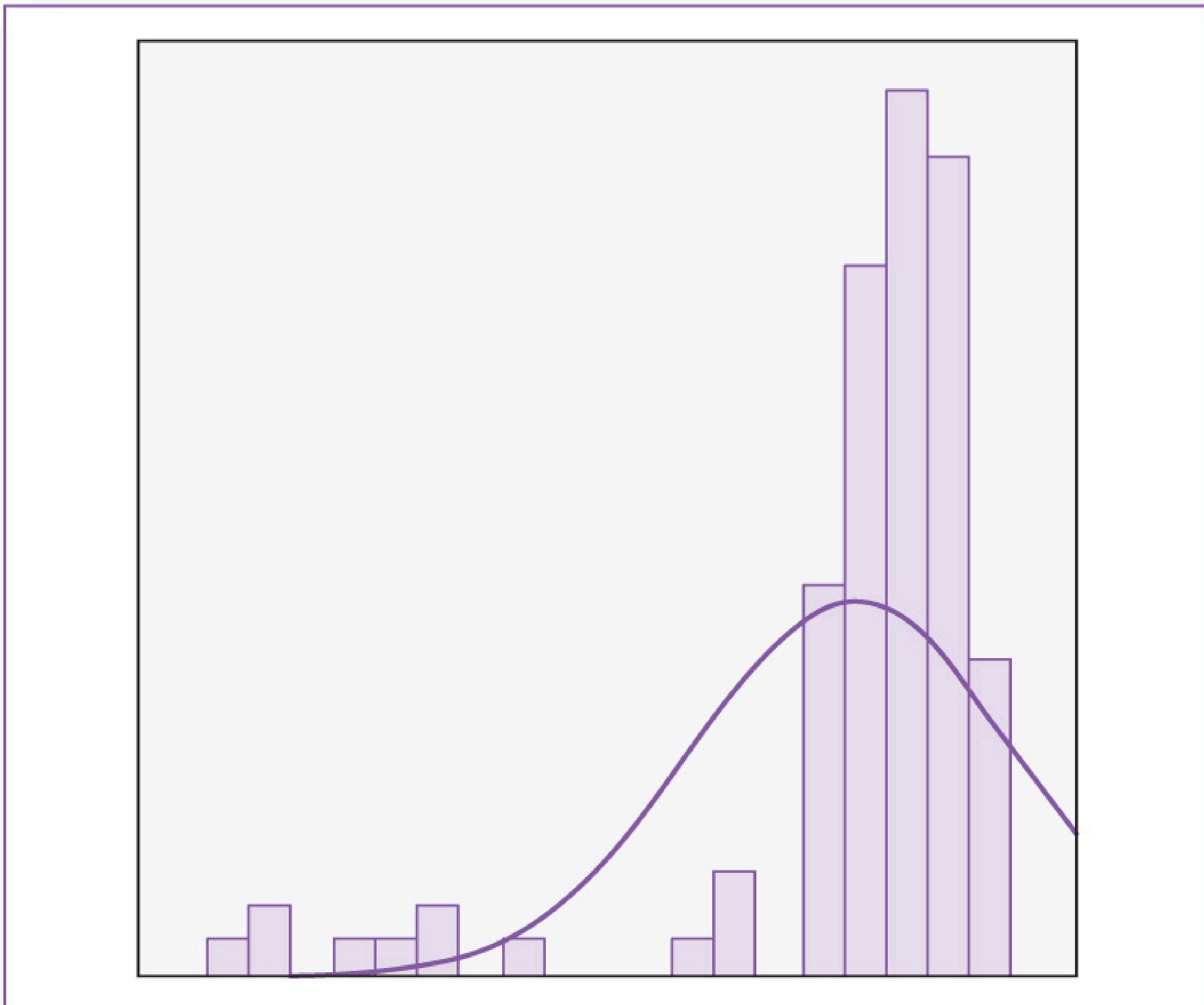


Figure 5.9 Bimodal distribution.

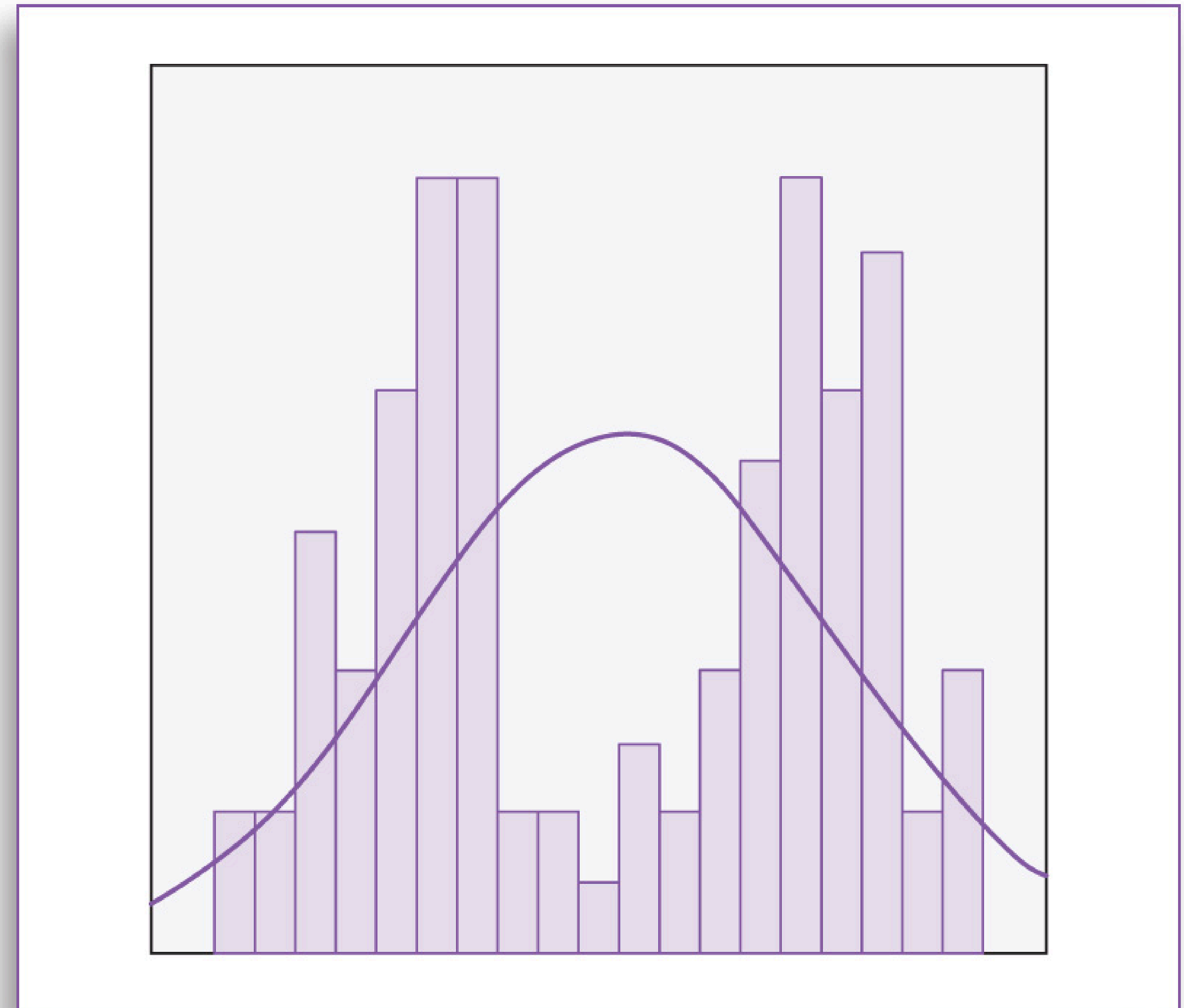
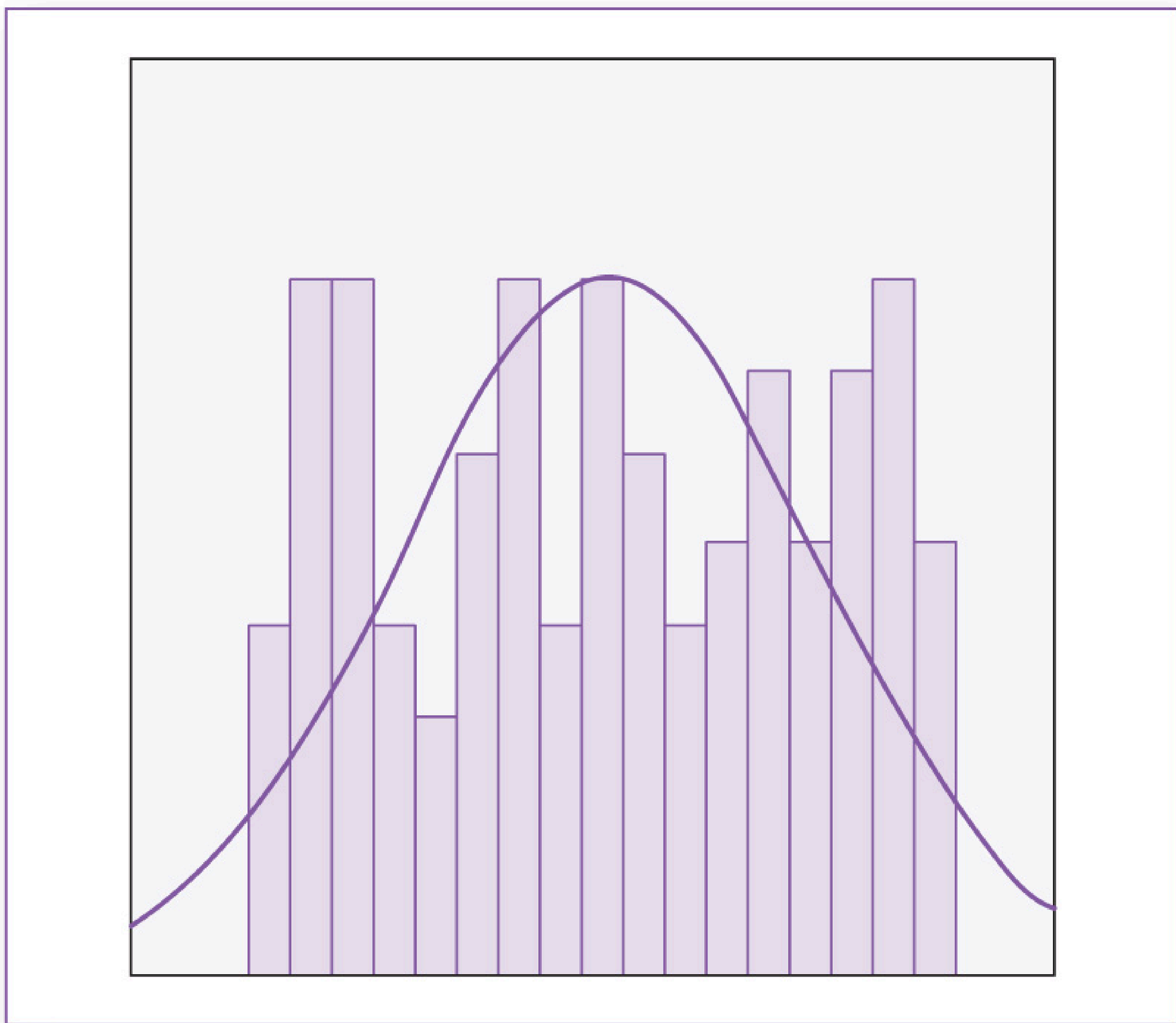


Figure 5.10 Flat distribution.

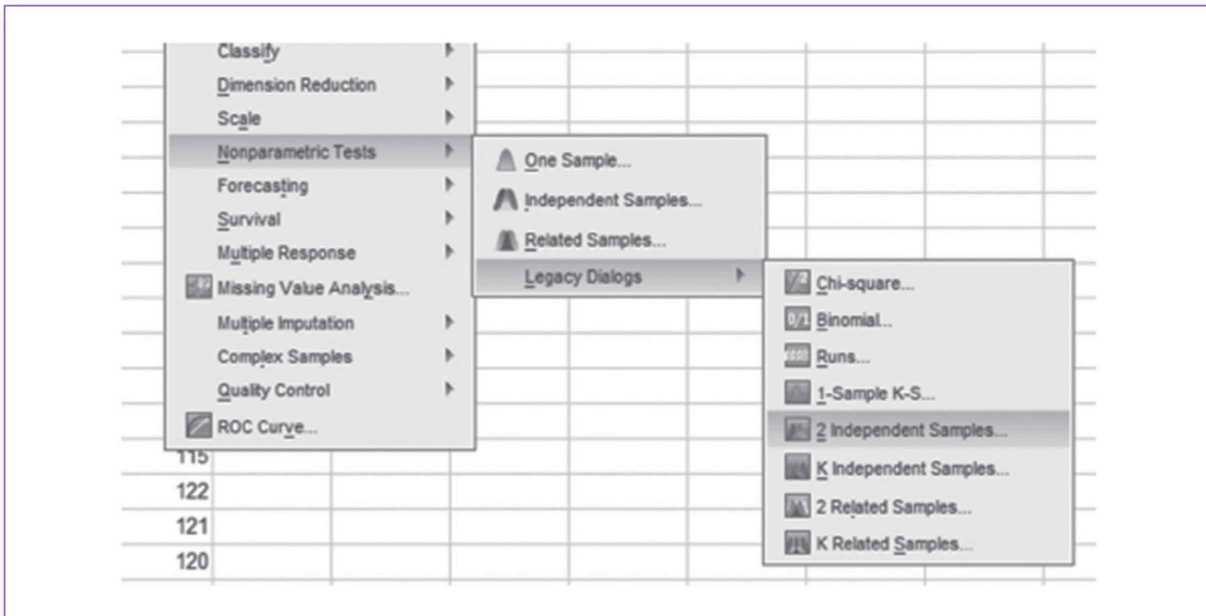


Test Run

For exemplary purposes, we will run the Mann-Whitney *U* test using the same data set (**Ch 05 - Example 01 - t Test and Mann-Whitney U Test.sav**), even though the data are normally distributed. This will enable us to compare the results of a *t* test with the results produced by the Mann-Whitney *U* test.

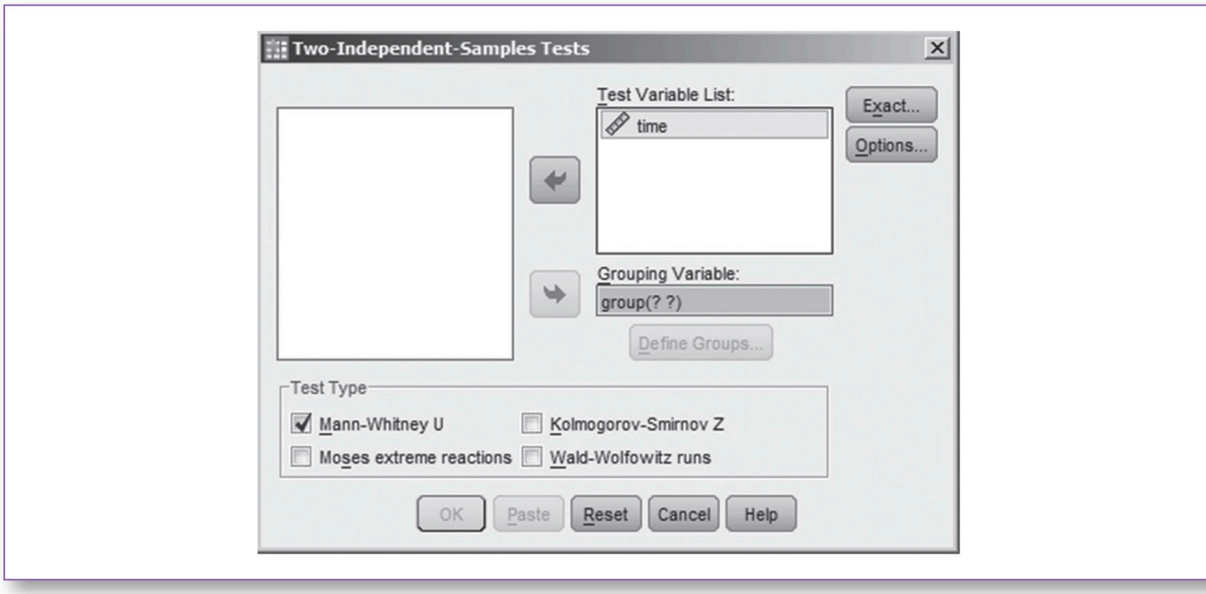
1. On the main screen, click on *Analyze*, *Nonparametric Tests*, *Legacy Dialogs*, *2 Independent Samples* (Figure 5.11).

Figure 5.11 Ordering the Mann-Whitney U test: click on Analyze, Nonparametric Tests, Legacy Dialogs, 2 Independent Samples.



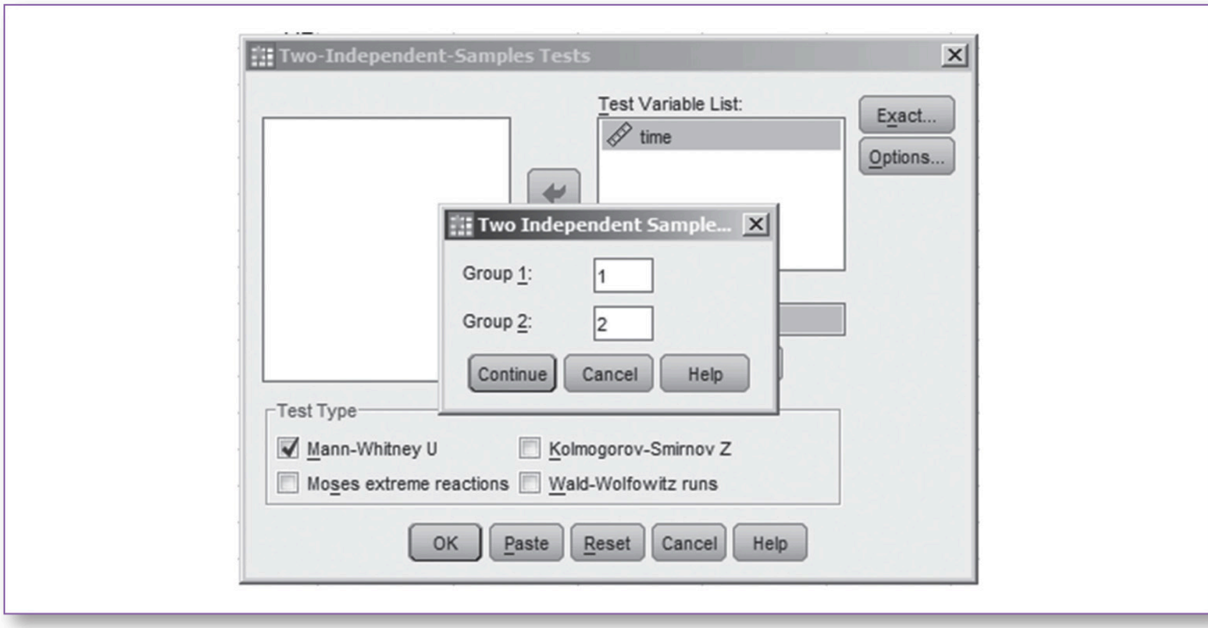
2. In the *Two-Independent-Samples Tests* window, move *time* to the *Test Variable List* panel.
3. Move *group* to the *Grouping Variable* box (Figure 5.12).

Figure 5.12 In the Two-Independent-Samples Tests window, move time to the Test Variable List panel, and move group to the Grouping Variable box.



4. Click on *group(? ?)*, then click on *Define Groups*.
5. In the *Two Independent Samples* dialog box, for *Group 1*, enter 1; for *Group 2*, enter 2 (this is because we defined Text as 1 and Text with illustrations as 2) (Figure 5.13).
6. Click *Continue*; this will close this dialog box.

Figure 5.13 In the Two Independent Samples dialog box, for Group 1, enter 1; for Group 2, enter 2.



7. In the *Two-Independent-Samples Tests* window, click on *OK*.

Results

The Mann-Whitney *U* test result is found in the *Test Statistics* table (Table 5.4); the Asymp. Sig. (2-tailed) statistic rendered a *p* value of .020; since this is less than α (.05), we would conclude that there is a statistically significant difference between the performances of the assembly instructions.

Table 5.4 Mann-Whitney U Test p Value = .020

Table 5.4 Mann-Whitney *U* Test *p* Value = .020

Test Statistics[a]

	time
Mann-Whitney U	415.000
Wilcoxon W	1045.000
Z	-2.322
Asymp. Sig. (2-tailed)	.020

a. Grouping Variable: group

Referring back, remember that the *t* test produced a *p* value of .015. The difference in this *p* values is due to the internal transformations that the Mann-Whitney *U* test conducts on the data. If one or more substantial violations are detected when running the pretest checklist for the *t* test, the Mann-Whitney *U* test is considered a viable alternative.

Good Common Sense

Clearly, it is essential that you comprehend key figures of the statistical reports you order, but when it comes to using these findings to make decisions in the real world, other considerations must be taken into account.

For example, suppose you measured the results of a new treatment to improve motivation using a 100-point scale (0 = low motivation, 100 = high motivation), wherein Group 1, the control group, received the usual treatment and Group 2 received the new treatment. Your analysis revealed that μ(Group 1) = 88.1, μ(Group 2) = 88.3, and $p = .01$. Your first reaction might be to leap to the conclusion that since $p = .01$, and this is less than the α level of .05, Group 2 statistically significantly outperformed Group 1, so we should adopt the treatment that was used for Group 2. Statistically speaking, you would be right, but consider some other real-world factors: Group 2 outperformed Group 1 by (only) .2 points; one might wonder if, in the real world, a .2-point difference is of any *practical* significance—can anyone really detect the difference between outcomes of 88.1 and 88.3?

Also, scarce resources may be an issue; the intervention used for Group 2 might have been very costly, time-consuming, labor intensive, or complex to carry out—it would be reasonable to consider if the cost, time, labor, and inconvenience involved are really worth the .2-point improvement.

Another concern might be the initial goal of the intervention: If the goal was to raise the score to at least 80, then clearly both groups achieved the desired effect—in this light, the innovative treatment would seem less impressive, especially if it is more complex, time-consuming, or costly to implement.

The lesson at hand is that when interpreting statistical findings, it is important that you not only tend to the numbers but also mindfully comprehend that those numbers are only a part of the picture when it comes to making *practical* decisions in the real world.

Key Concepts

- *t* test
- Pretest checklist
    - Normality
    - *n*
    - Homogeneity of variance
- α
- *p*
- Hypothesis resolution
- Documenting results
- Mann-Whitney *U* test
- Type I (α) error
- Type II (β) error
- Good common sense

Practice Exercises

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

Exercise 5.1

You want to determine if meditation can reduce resting pulse rate. Participants were recruited and randomly assigned to one of two groups: Members of Group 1 (the control group) did not meditate; members of Group 2 (the treatment group) meditated for 30 minutes per day on Mondays, Wednesdays, and Fridays over the course of 2 weeks. At the end of the study, you gathered the resting pulse rates for each participant.

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

Codebook

Variable: group
Definition: Meditation group assignment
Type: Categorical (1 = No meditation, 2 = Meditates 3 days)
Variable: pulse
Definition: Pulse rate measured in number of 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 *t* test and document your findings (*n*s, means, and significance [*p* value]).
d. Write an abstract up to 200 words detailing a summary of the study, the *t* test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 05 - Exercise 01B.sav**.

Exercise 5.2

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 assigned to the mentor group (and will be given mentors) or to the control group (no mentor) on an alternating basis. Those in the mentor group are to meet their mentors in person once a week at a time of their choosing. All freshmen, in both groups, agree to submit their transcripts at the conclusion of the term.

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

Codebook

Variable: group
Definition: Mentor group assignment
Type: Categorical (1 = No mentor, 2 = In-person 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 *t* test and document your findings (*n*s, means, and significance [*p* value]).
d. Write an abstract up to 200 words detailing a summary of the study, the *t* test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 05 - Exercise 02B.sav**.

Exercise 5.3

Clinicians at a nursing home facility want to see if giving residents plants to tend to will help lower depression. To test this idea, the residents are randomly assigned to two groups: Each member of the treatment group will be issued a small bamboo plant along with a card detailing care instructions; members of the control group will receive no plant. 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 05 - Exercise 03A.sav**

Codebook

Variable: group

Definition: Plant group assignment

Type: Categorical (1 = No plant, 2 = Bamboo)

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 *t* test and document your findings (*n*s, means, and significance [*p* value]).

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

Repeat this exercise using data set **Ch 05 - Exercise 03B.sav**.

Exercise 5.4

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

Data set: **Ch 05 - Exercise 04A.sav**

Codebook

Variable: group

Definition: Chocolate dosage group assignment

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

Variable: mood

Definition: Score on 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 *t* test and document your findings (*n*s, means, and significance [*p* value]).

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

Repeat this exercise using data set **Ch 05 - Exercise 04B.sav**.

Exercise 5.5

During flu season, the administrators at a walk-in health clinic want to determine if providing patients with a pamphlet will increase their receptivity to flu shots. Once escorted to the exam room, patients who had odd-numbered service tickets were given a flu shot information pamphlet; patients with even-numbered tickets were given nothing. At the end of the day, patients' charts were reviewed, and two entries were made in the database: total number of flu shots given to patients who received pamphlets and total number of flu shots given to patients who were not given flu shot pamphlets.

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

Codebook

Variable: group

Definition: Flu media group assignment

Type: Categorical (1 = Nothing, 2 = Flu shot pamphlet)

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 *t* test and document your findings (*n*s, means, and significance [*p* value]).

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

Repeat this exercise using data set **Ch 05 - Exercise 5B.sav**.

Exercise 5.6

You want to determine if watching a video of a comedy with a laugh track is more enjoyable than watching without it. Subjects will be recruited and randomly assigned to one of two groups: Those in the control group will watch the video without the laugh track, and those assigned to the treatment group will watch the same video with the sound of a 50-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 05 - Exercise 06A.sav**

Codebook

Variable: group

Definition: Laugh track group assignment

Type: Categorical (1 = No laugh track, 2 = Laugh track at 50)

Variable: enjoy

Definition: Enjoyment of video

Type: Continuous (1 = not very enjoyable, 5 = very enjoyable)

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 *t* test and document your findings (*n*s, means, and significance [*p* value]).

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

Repeat this exercise using data set **Ch 05 - Exercise 06B.sav**.

Exercise 5.7

In an effort to determine the effectiveness of light therapy to alleviate depression, you recruit a group of subjects who have been diagnosed as depressed. The subjects are randomly assigned to one of two groups: The control group will receive no light therapy, and the treatment group will get light therapy for 1 hour on even-numbered days over the course of 1 month, at which time 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 07A.sav**

Codebook

Variable: group
Definition: Light therapy group assignment
Type: Categorical (1 = No light therapy, 2 = Light therapy: even days)
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 *t* test and document your findings (*n*s, means, and significance [*p* value]).
d. Write an abstract up to 200 words detailing a summary of the study, the *t* test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 05 - Exercise 07B.sav**.

Exercise 5.8

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 two groups: For 30 days, members of the experimental group will walk from 7:00 to 7:30 a.m., Monday through Friday; members of the control group will do no walking. At the conclusion of the study, each subject 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 08A.sav**

Codebook

Variable: group

Definition: Walking group assignment

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

Variable: energy

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 *t* test and document your findings (*n*s, means, and significance [*p* value]).

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

Repeat this exercise using data set **Ch 05 - Exercise 08B.sav**.

Exercise 5.9

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 subjects to read a 1,000-word essay. Half of the subjects will be randomly assigned to the control group; they will read the essay using regular room lighting. The other half of the subjects will read the same essay using the Acme reading lamp.

Data set: **Ch 05 - Exercise 09A.sav**

Codebook

Variable: group
Definition: Lighting group assignment
Type: Categorical (1 = Room lighting, 2 = Acme lamp)
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 *t* test and document your findings (*n*s, means, and significance [*p* value]).
d. Write an abstract up to 200 words detailing a summary of the study, the *t* test results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set **Ch 05 - Exercise 09B.sav**.

Exercise 5.10

You want to find out if music enhances problem solving. Subjects will be recruited and randomly assigned to one of two groups: Those in the control group will be given a standard 100-piece jigsaw puzzle to solve in a quiet room. Participants in the experimental group 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. You will record the time it takes for each person to complete the puzzle.

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

Codebook

Variable: group

Definition: Music group assignment

Type: Categorical (1 = No music, 2 = Music at 30 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 *t* test and document your findings (*n*s, means, and significance [*p* value]).

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

Repeat this exercise using data set **Ch 05 - Exercise 10B.sav**.

https://doi.org/10.4135/9781071878910