Chapter 9

Comparing Two Means: The t-test

This chapter focuses on evaluating differences between mean values from two different groups. Oftentimes, researchers wish to know whether groups are similar or different from each other. For example, do college students watch more television than the average adult? Are men more jealous than women? Do those given a lecture on the benefits of exercise actually increase their exercise behaviors compared to those who didn't receive the lecture? The research situation is one with a continuous (i.e., scale measured) dependent variable, and we are interested in comparing two group mean values. All of these examples contain two groups that can be compared: college students vs. the average adult; men vs. women; individuals given a lecture vs. those who did not. To assess who watches television more, has greater jealousy, and exercises more, a *t*-test can be performed.

We start with the simplest application of the *t*-test, the one-sample *t*-test. The one-sample *t*-test is designed to compare a *sample* mean to a (a) population mean, or (b) pre-determined value. *Case Study 9.1* compares college student television viewing time to the general adult population, while *Case Study 9.2* evaluates gains in knowledge on a sexually transmitted disease (STD) test. We next move to comparing mean values from independent samples, where a *t*-test for independent samples is used. Sometimes called Student's t-test (*Student* was the pseudonym that Henry Gossett used to publish because his employers at Guinness Brewery didn't want competitors to know they were using statistics to assist in beer making), it is used if we want to assess whether the mean score from one group of individuals differs from the mean score of another group. *Case Study 9.3* evaluates mean gender differences on sexual jealousy, while *Case Study 9.4* is an experiment intended to compare exercise behavior between an intervention group and a control group. Also in this chapter, we cover the assumptions of the *t*-test. The end of the chapter provides *t*-test applications using both IBM SPSS and SAS.

Case Study 9.1: Who Watches more Television – College Students or the General Adult Population?

Over the past 40 years, television has become a common form of household entertainment. In the average American household, a television is on for 8 hours a day, and the average American adult watches for 5 hours per day. College students sometimes are stereotyped as watching a lot of television. But do they watch television more than the average adult, less than the average adult, or do they watch the same amount?

In this section, we consider the *t*-test as applied to a single sample of individuals (in this case, college students), and how much television they watch on a given day. This type of *t*-test is called a *one-sample t-test* since a single sample mean is compared to a preexisting population value. To conduct this study, we sample 10 college students who live in a dorm and ask how much television they watched in the past 24-hours. Regardless of the stereotype, we believe that college students watch less than 5 hours of television a day. Therefore, the research hypothesis (H_1) is that college students watch less television per day than the average American adult. The null hypothesis (H_0) is that the sample and population mean will be equal (H_0: $M = \mu$; where M is the sample mean and μ is a population mean). Data from these 10 students are represented in Table 9.1.

Summary Statistics: Mean, Variance, and Standard Deviation

As noted in earlier chapters, an evaluation of basic descriptive statistics first should be performed. For our sample of 10 college students, the mean, variance, and standard deviation values can be calculated. Table 9.2 contains the summary calculation components.

Mean. The mean value for television viewing in a 24 hour period is 3.90: 39/10 = 3.90. On average, the 10 college students watch 3.9 hours of television per day. We already know that the population mean is 5, so it appears that at least descriptively the sample mean is lower than the population mean.

Mean for College Students: $M = \sum X_i/n = 39/10 = 3.90$

Variance. The variance is the level of dispersion in a sample and is calculated by summing the squared deviations and dividing by *n*-1. For our current sample, the variance is 1.21.

Variance for College Students: $$s^2 = \frac{\sum (X_i - M)^2}{n-1} = \frac{10.90}{9} = 1.21$$

Standard deviation. The standard deviation is the average spread of scores in a sample, and is calculated by taking the square root of the variance. For the 10 college students, the standard deviation is 1.10.

Standard Deviation for College Students: $$s = \sqrt{s^2} = \sqrt{1.21} = 1.10$$

Descriptive Data Plots

Viewing a histogram of the data provides a sense of the distribution of the data. A histogram of student television viewing in a 24-hour period is presented in Figure 9.1. Here, even with a small sample of 10 cases, the data appear normally distributed.

The One-Sample t-test

After evaluating descriptive statistics and a histogram, we proceed to test the null hypothesis. Whether our college sample differs from the average American adult in television watching will be assessed using the one-sample *t*-test.

The one-sample *t*-test is calculated by taking the difference between the sample mean and the population mean or preexisting value, and dividing it by the standard error of the sample population (the standard error was covered in Chapter 2). By doing this, we are standardizing the mean difference, essentially converting it to a z-score. Formula 9.1 for the one-sample *t*-test is noted below:

$$t = \frac{M - \mu}{\sqrt{\frac{s^2}{n}}}$$ *Formula 9.1: One-Sample t-test*

The *t* is the symbol for the *t*-test value. *M* is the mean for the sample. The population mean is represented by μ (called Mu). If we are comparing our sample mean to a preexisting value, we would replace μ with that number. S^2 is the sample variance, and *n* is the sample size.

One-sample *t*-tests can be positive or negative, and have no upper or lower range. T-test values can be positive or negative, just like a Pearson correlation. The positive or negative value reflects whether the sample mean is larger or smaller than the population mean. Larger values indicate a greater difference. Because the resulting *t*-test value is similar to a z-score, with larger samples a *t*-test value over +/- 1.96 is typically significant at the $p < .05$ level (two-tailed). There is no upper or lower bound for the *t*-test, so resulting values can take on any value.

Numerator

The numerator of the one-sample *t*-test is the sample mean subtracted from the population mean μ. As noted earlier, the population mean μ can be replaced with a preexisting value depending on the research problem.

Denominator

The denominator of the one-sample *t*-test is the sample standard error. The calculation takes the sample variance and divides it by the sample size, and the square root is taken.

Steps in calculating the Denominator of the One-sample t-test

1) Calculate the variance from the sample data

2) Divide the variance by the sample *n*, then take the square root

At this point, the one-sample *t*-test can be calculated, dividing the numerator by the denominator.

Assessing the Statistical Significance of the One-Sample t-test

Once the one-sample *t*-test has been calculated, we assess whether the value – which represents the standardized difference between the sample mean and population mean -- exceeds what is expected by chance. This is done by finding a critical value at which we would accept the null hypothesis given a specific probability level.

To find a *t* critical value, we do three things. First, we need to adopt an alpha level – usually a .05 significance level will suffice, although more stringent alpha levels such as .01 are sometimes used. Second, we need to adopt a one- or two-tail test of significance based on how the null hypothesis is stated. Usually this will be a two-tail test, but in some circumstances with the one-sample *t*-test you might find a one-tail test applied. How the null hypothesis is stated will guide your choice of a one- or two-tailed test. Third, *degrees of freedom* (*df*) are calculated. Formula 9.2 is the formula for the one-sample *t*-test degrees of freedom:

$$N - 1 = \text{Degrees of Freedom} \qquad \textit{Formula 9.2}$$

Once these steps have been completed, Appendix T is used to find the *t* critical value. In Appendix T, find the appropriate row using the degrees of freedom, then find the appropriate column based on the selected probability (e.g., .05, two-tail test). Once the column is located, the *t* critical value is found. If the *t*-test value exceeds the *t* critical value taken from Appendix T, we reject the null hypothesis and conclude that the mean difference is beyond what would be expected by our probability level.

6 Steps for Determining the Significance of the One-Sample t-test

1) **Adopt a statistical significance level (usually .05)**

2) **Choose a one- or two-tailed test based on the null hypothesis**

3) **Calculate the degrees of freedom using *N* – 1**

4) **Use *Appendix T* to find the *t* critical value**. Note that Appendix T lists the *t* critical values as positive, but these values would also apply to negative associations. For example, a *t* critical value in Appendix T of 1.56 also reflects the value of -1.56.

5) **Ask whether the *t*-test value calculated for the study exceeds the *t* critical value**. If the study *t*-test value exceeds the *t* critical value – be it either negative or positive – then the finding is statistically significant and we reject the null hypothesis. The resulting difference between the sample and population mean exceeds what would be expected by our probability level. If the study *t* value does not exceed the critical *t* value, then the mean difference is not statistically significant and the null hypothesis is accepted.

6) **Interpret the results.** Besides reporting whether the resulting *t*-test value is significant, interpreting results should also include the direction of the finding. If the finding is non-significant, we should acknowledge confirmation of the null hypothesis. For example, stating the college student sample differs from the general population in regard to television viewing is only a partial interpretation. An appropriate interpretation would include whether college students watched more or less television than the average adult.

An alternative way to assess the significance of the one-sample *t*-test is to use a statistical program such as IBM SPSS or SAS. No *t* critical value is required, nor do we need Appendix T, because the programs provide the exact probability of the *t*-test value given the null hypothesis is true. If the exact probability is less than .05, then we have a significant finding; the probability that our results are due to chance is less than 5%.

Effect Strength

Once the one-sample *t*-test statistic is derived and assessed, it may be converted to *r* as a measure of effect strength. As covered in Chapter 5 on effect sizes, an *r* value of .10 indicates a small effect, meaning that group membership has a minimal influence on the dependent variable or outcome variable. An *r* value of .30 indicates a medium effect, and values .50 indicate a large effect. Formula 9.3 for converting *t* to *r* is presented below:

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

Formula 9.3 Conversion of t to r for Effect Size

Here, the square root is taken of the squared *t*-test value divided by the sum of the squared *t*-test value and degrees of freedom. By converting *t* to *r*, we now have a familiar effect size to evaluate the impact or effect of group membership on the resulting mean difference between the sample and the population on the dependent variable or outcome variable.

Applying the One-Sample t-test to Case Study 9.1: Who Watches more Television – College Students or the General Adult Population?

Returning to our television watching example, the null hypothesis is that the mean hours watching television in a 24-hour period for college students and the average American adult will be the same. The research hypothesis is that college students will watch less television in a 24-hour period than the average adult.

Steps One-Sample t-test formula for Television Viewing Study

1) $t = \frac{M - \mu}{\sqrt{\frac{s^2}{n}}}$

2) $t = \frac{3.90 - 5.00}{\sqrt{\frac{1.21}{10}}}$

3) $t = \frac{-1.10}{\sqrt{0.12}}$

4) $t = \frac{-1.10}{0.348}$ = -3.16

Numerator

The mean of 3.9 for college students is subtracted from the average adult mean of 5.0. In our example, the numerator is 3.9 – 5.0 = -1.10.

Denominator

The denominator for the one-sample *t*-test is the standard error of the sample mean. The denominator is calculated by taking the sample variance of 1.21 and dividing by the sample size (N = 10), producing a value of 0.212: 1.21/10 = 0.121. The square root of this value is then taken, which gives us a standard error of 0.348: $\sqrt{0.121}$ = 0.348. The numerator of -1.10 is divided by 0.348, producing a one-sample *t*-test statistic of -3.16: -1.10/0.348 = -3.16.

Assessing the Statistical Significance of Case Study 9.1

We have a *t*-test value of -3.16. Let's follow the 6 steps to assess whether this value is significant.

Steps for Assessing the Significance of the One-Sample t-test

1) Adopt a statistical significance level (usually .05)

- We will adopt the .05 level of significance

2) Choose a one- or two-tailed test based on the null hypothesis

- We will choose a two-tail test based on our null hypothesis

3) Calculate the degrees of freedom using *N - 1*

- For the current example, the degrees of freedom (N - 1) is 9: 10 – 1 = 9

4) Use Appendix T to find the *t* critical value

- Using a two-tailed test with a .05 level of significance, Appendix T shows a *t* critical value of +/- 2.26.

5) Ask whether the *t*-test value calculated for the study exceeds the critical *t* value.

-Does the *t*-test value of -3.16 exceed the *t* critical value of +/- 2.26? It does, and we can conclude that the number of hours college students watch television is significantly smaller compared to individuals in the average American household. The chance of this mean difference occurring by chance is less that 5%.

6) Interpret the results

- Besides reporting whether the resulting *t*-test value is significant, interpreting results should also include the direction of the finding or confirmation of the null hypothesis. For example, stating that our college student sample differs from the general population in regard to television viewing is only a partial interpretation. An appropriate interpretation would include whether college students watched more or less television than the average adult.

The second way to make our evaluation is to analyze the data using IBM SPSS or SAS. Using these data, SPSS tells us that the probability of getting a *t*-test value of -3.16 is equal to

.012 (SAS is even more specific, reporting $p = .0115$). In other words, the probably of getting a *t*-test value of -3.16 given the null hypothesis is true is .012, which exceeds our selected probability level of $p < .05$. These analyses are presented in our section at the end of the chapter on statistical analyses in SPSS and SAS for this chapter.

Effect strength

Using Formula 9.3, we produce an effect size which indicates how big an effect being a member of the sample has on the resulting mean difference. In other words, how much of an effect (small, medium, or large) does being a college student have on the resulting difference between the sample mean and population mean? Using the 4 steps below, the resulting effect size based on *r* is 0.725, a large effect. Being a college student has a large effect on the resulting difference between the sample and population means.

Steps	Formula
1)	$r = \sqrt{\frac{t^2}{t^2 + df}}$
2)	$r = \sqrt{\frac{-3.16^2}{-3.16^2 + 9}}$
3)	$r = \sqrt{\frac{9.9856}{9.9856 + 9}}$
4)	$r = 0.725$

Example Write-up

A study was conducted to evaluate television viewing over a 24-hour period by college students compared to the average American adult. The research hypothesis was that college students would watch less television than the average adult. The null hypothesis was that college students would watch the same amount television as the average adult, H_0: $M = \mu$. *From a* *national study of television viewing, it was ascertained that television viewing for the average adult was 5-hours per day. Ten college students participated in the study, and were asked to document the amount of television they watched in the past 24-hours. The overall sample mean was 3.90 (sd = 1.10). A one-sample t-test was performed to compare the college sample mean to the overall population mean of 5.0. The mean difference was found to be statistically significant beyond the .05 level of significance, $t(9) = -3.16$, $p < .05$, and the effect size was large ($r = 0.725$). Overall, we conclude that college students watch less television in a 24-hour period ($M = 3.90$) than the average adult, and reject the null hypothesis.*

Case Study 9.2: Young Adult knowledge gain in sexually transmitted disease (STD) information

One of your textbook authors recently completed a study to increase knowledge of sexually transmitted diseases (STDs) in young adults. STDs are diseases such as gonorrhea, Chlamydia, herpes, and HIV (just to name a few). In 2008 the Centers for Disease Control and Prevention estimated that in the United States alone, there were 19 million new STD cases, and almost half were found in young adults ages 15-24 (Centers for Disease Control and Prevention, 2009).

The research hypothesis (H_1) was that individuals over a one-month period would show a gain in STD knowledge measured as the number of correct items on a 20-item true/false test after being exposed to information regarding STDs and their spread. The null hypothesis was that the change in STD knowledge scores would be equal to zero (H_0: $M = 0$). Note here that we use the value zero (0) in the null hypothesis statement as a “pre-determined value” since no change in STD knowledge scores would result in a gain score of zero (0).

Here, we will not provide the raw data, nor do we need it. The *t*-test requires only the mean value, variance or standard deviation, and sample size for calculations. Based on a sample of 20 individuals who were given an STD information lecture, the mean *gain* in STD knowledge over a one-month period was 3.5, with a variance of 5.0 and standard deviation of 2.236. On average, students increased their STD knowledge by 3.5 points. Is this gain of 3.5 significantly different from a predetermined value of 0, indicating no gain?

Applying the One-Sample t-test to Case Study 9.2: Young Adult knowledge gain in sexually transmitted disease (STD) information

For this example, we start by first presenting the one-sample t-test formula (see below) with sample values included:

Steps *One-Sample t-test formula for STD Knowledge Gain Study*

1) $$t = \frac{\overline{X} - \mu}{\sqrt{\frac{s^2}{n}}}$$

2) $$t = \frac{3.5 - 0}{\sqrt{\frac{5.0}{20}}}$$

3) $$t = \frac{3.5}{\sqrt{0.25}}$$

4) $$t = \frac{3.5}{0.50} = 7.0$$

For the numerator, the sample mean of 3.5 is subtracted from 0 (the predetermined value indicating no gain in knowledge). The resulting value for the numerator is 3.5. For the denominator, the sample variance of 5.0 is divided by the sample size (20), which derives a value

0.25: 5.0/20= 0.25. We next take the square root of this value which gives us the denominator of 0.50: $\sqrt{0.25} = 0.50$. The numerator of 3.5 is now divided by the denominator of 0.50, producing a one-sample *t*-test statistic of 7.0: 3.5/0.50 = 7.0.

Assessing the Statistical Significance of Case Study 9.2

Our *t*-test value is 7.0. Next, we evaluate the significance of the one-sample *t*-test using the following steps.

6 Steps for Assessing the Significance of the One-Sample t-test

1) **Adopt a statistical significance level (usually .05)**

   - We will adopt the .05 level of significance

2) **Choose a one- or two-tailed test based on the null hypothesis**

   - We will choose a two-tail test based on our null hypothesis

3) **Calculate the degrees of freedom using *N - 1***

   - For the current example, the degrees of freedom (N - 1) is 19: 20 – 1 = 19

4) **Use Appendix T to find the *t* critical value**

   - Using a two-tailed test with a .05 level of significance, Appendix T shows a critical *t* value of +/- 2.09.

5) **Ask whether the *t*-test value calculated for the study exceeds the *t* critical value.**

   - Does our *t*-test value of 7.0 exceed the critical value of +/- 2.09? It does, and we conclude that young adults gain in STD knowledge was significantly different from zero. The mean difference exceeded our probability level of $p < .05$.

6) **Interpret the results.**

- Young adults show positive change in STD knowledge over a 1-month period, with a mean gain in correct answers on a 20-item test of 3.5. We therefore reject the null hypothesis that H_0: $M = \mu$.

Because we do not have the raw data, we cannot perform the analysis in IBM SPSS or SAS. Most statistical programs require the raw data to perform the required statistical tests.

Effect strength

Using Formula 9.3, the resulting effect size based on *r* is 0.85, a large effect. Attending the STD lecture has a large effect on the improvement of scores on an STD knowledge test.

Step	Formula
1)	$r = \sqrt{\frac{t^2}{t^2 + df}}$
2)	$r = \sqrt{\frac{7.0^2}{7.0^2 + 19}}$
3)	$r = \sqrt{\frac{49.0}{49.0 + 19}}$
4)	$r = 0.85$

Example Write-up

A study was conducted to evaluate gains in sexually transmitted disease (STD) knowledge in a small sample of college students who were exposed to information on STDs and their spread. The research hypothesis is that the information presented to students would increase their STD knowledge. The null hypothesis is that STD knowledge gain would be equal to zero, H_0: $M = 0$. *Twenty young adults attending a local university participated in the study, and were asked to complete an STD knowledge scale at the beginning of the study, and then again 1-month after they had received information regarding STDs. Overall, a mean gain of 3.5 points on the*

STD knowledge scale was noted for the sample. A one-sample t-test was used to compare the gain in STD knowledge to a value of zero (indicating no gain). The average gain in STD knowledge by the sample was found to significantly exceed zero, t(19) = 7.0, p < .05, with a large effect (r = .85). We therefore reject the null hypothesis, and conclude the information presented to the young adults increased their STD knowledge. Overall, young adults showed a significant gain in STD knowledge across the one-month period, with an average gain of 3.5-points on a 20-item True/False test.

Comparing Two Sample Means: The Independent Samples t-test

The *t*-test for independent samples is used when there are two sample means to be compared. In Case Studies 9.1 and 9.2, only one sample mean was available and was compared to a population mean or preexisting value. In this section, we cover the situation where two samples means are compared. *Case Study 9.3* addresses gender differences in sexual jealousy, testing whether men report higher levels of sexual jealousy than women, while *Case Study 9.4* is an intervention study to increase exercise behavior. In between Case Studies 9.3 and 9.4, we cover the data assumptions of the *t*-test, and illustrate assessment of these assumptions for both examples.

Case Study 9.4: Gender Differences in Sexual Jealousy

As we first noted in Chapter 4, jealousy is a common and recurring theme in close relationships. Most individuals admit they have been jealous at some point in their lives. Who is more likely to be jealous, men or women? Based on prior research, we might hypothesize that men will be more jealous than women, especially if we focus on *sexual jealousy* where one's partner is suspected of having a sexual affair. Thus, we derive a research hypothesis that reflects this belief, *H_1: Males will express higher levels of sexual jealousy than females*. The null hypothesis will reflect that male and female sexual jealousy will be the same, so we write the null hypothesis as H_0: $\mu_1 = \mu_2$, with μ_1 being the jealousy mean for males, and μ_2 being the jealousy mean for females.

To conduct such a study, we ask 30 individuals (15 male and 15 female) from an Introductory Psychology course to complete a short measurement scale on sexual jealousy. Participants are given a 10 item sexual jealousy measure with items focusing on sexual infidelity situations, and are asked to rate each situation on a 5-point scale ranging from 1 (not jealous at all) to 5 (very jealous). An example situation is, "You see your partner flirting with another." Higher scores indicate greater sexual jealousy, and the scale is assumed to be a reliable measure. Table 9.3 contains data for the 30 individuals. The data are listed in two columns with one column for each gender.

Summary Statistics: Means, Variances, and Standard Deviations

The mean, variance, and standard deviation are derived for each gender, and are noted in Table 9.4 and discussed below.

Mean. In our example, the mean for Males is 3.6: 54/15 = 3.6, and for Females the mean is 2.8: 42/15 = 2.8. Based on these means, it appears that males have higher levels of sexual jealousy than females.

Mean for Male Sexual Jealousy: $M = \sum X_i/n = 54/15 = 3.6$

Mean for Female Sexual Jealousy: $M = \sum X_i/n = 42/15 = 2.8$

Variance. For Males, the variance is 0.83: 11.6/14 = 0.83. For Females, the variance is 1.17: 16.4/14 = 1.17. A summary of the calculations is below, and Table 9.4 contains the calculation steps for the variance for our data.

Variance for Male Sexual Jealousy: $s^2 = \frac{\sum(X_i - M)^2}{n-1} = \frac{11.6}{14} = 0.83$

Variance for Female Sexual Jealousy: $s^2 = \frac{\sum(X_i - M)^2}{n-1} = \frac{16.4}{14} = 1.17$

Standard deviation. For Males, the standard deviation is 0.91: $\sqrt{0.83} = 0.91$. For Females, the standard deviation is 1.08: $\sqrt{1.17} = 1.08$.

Standard Deviation for Male Sexual Jealousy: $s = \sqrt{s^2} = \sqrt{0.83} = 0.91$

Standard Deviation for Female Sexual Jealousy: $s = \sqrt{s^2} = \sqrt{1.17} = 1.08$

Descriptive Data Plots

Besides looking at mean and standard deviation values, plotting the data within each group visually shows the distribution of the data. A good starting point is simply plotting the mean values for each group in a bar chart (see Figure 9.2). Bar charts help illustrate the means and how they differ. Viewing histograms of the data from each group also provide a visual display of the distributions. The histogram for Male sexual jealousy (see Figure 9.3) appears bell-shaped, as does the histogram for Females (see Figure 9.4). Later we will see how histograms are used to assess the assumption of normality for the *t*-test.

The Independent Samples t-test

The independent samples *t*-test is calculated by taking the difference between the group means and dividing by a standard error estimate. In essence, as with the one-sample *t*-test, we are standardizing the mean difference. The formula for the *t*-test is noted below in Formula 9.4.

$$t = \frac{M_1 - M_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Formula 9.4: Independent Samples t-test

As can be seen in Formula 9.4, the numerator of the independent samples *t*-test is the difference between the two sample mean values ($M_1 - M_2$), with the subscripts (1 and 2) indicating Group 1 and Group 2. The denominator pools the sample variances (s_p^2; which we will explain below), and creates a standard error estimate by dividing the pooled sample variances by the sample size in each group, summing these values, then taking the square root of the result. Both the numerator and denominator calculations are explained in greater detail below.

What does the resulting *t*-test value represent? First, we can simply think of the resulting value as the standardized difference between two means. The significance of this standardized difference is then assessed in a similar fashion to how a z-score is assessed. A second way to view the *t*-test is that the formula converts data for any two groups to a single value. Our null hypothesis expectation is that the means will be exactly equal and so the value of *t* will be 0.00. The value of *t* becomes larger when the data (i.e., the difference between the means obtained in the research) conform to the research hypothesis expectation rather than the null hypothesis expectation.

The independent samples *t*-test can be positive or negative, and has no upper or lower range. As with the one-sample *t*-test, the independent samples *t*-test can be positive or negative. This is solely due to the order of the mean values in the numerator (we suggest ordering your mean values based on how your null hypothesis is stated). In addition, *t*-test values can take on any value since there are no upper or lower bounds on the resulting value (unlike a correlation coefficient, which is bounded between -1 and +1).

Numerator

The *t*-test formula numerator is the difference between the two sample means. As a general rule, the mean values are placed in the same order as written in your null hypothesis.

Denominator

The denominator of the *t*-test formula is called the *standard error of the difference between the means*, which is a standard error estimate based on the group variances. In most instances[1], it uses a pooled variance value s_p^2, which is calculated using Formula 9.5.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$ *Formula 9.5: Pooled Variance*

In this pooled variance formula, the group variances are multiplied by *n* -1 and summed, then divided by the sum of the group sizes minus 2. This pooled variance creates a weighted average of the two group variances, summarized into a single pooled estimate.

Returning to the *t*-test denominator $\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$, once the pooled variance s_p^2 is calculated, we take s_p^2 and divide by the group sample sizes (*n*), and the two ratios are then summed. The square root of that sum is the standard error of the difference between the means.

Steps in calculating the Denominator of the Independent Samples t-test using the Pooled Variance approach:

1) **Calculate the pooled variance of both groups**
2) **Take the pooled variance and divide by the sample sizes for each group, add the resulting ratios, then take the square root**

---

[1] Note that later in this chapter when we addressing assumptions of the *t*-test, we will present an altered denominator approach for the *t*-test using a *separate variance estimation* approach.

At this point, the independent samples *t*-test can be calculated, dividing the numerator by the denominator.

Assessing the Significance of the Independent Samples t-test

Once a *t*-test value has been calculated, we need to assess whether it exceeds what is expected by chance. This is done by finding a critical value at which we would accept the null hypothesis given a specific probability level. If the *t*-test value exceeds the critical value, then its a significant finding, exceeding our adopted probability level.

To find a *t* critical value, we do three things. First, adopt an alpha level – usually a .05 significance level is appropriate, although a .01 level is sometimes used. Second, adopt a one- or two-tail test of significance based on the null hypothesis. Third, *degrees of freedom* (df) are calculated. Formula 9.6 is the formula for the independent samples *t*-test degrees of freedom:

$$N - 2 = \text{Degrees of Freedom} \qquad \textit{Formula 9.6}$$

Once we have made these decisions, Appendix T is used to find the *t* critical value. In Appendix T, find the appropriate row using the degrees of freedom, and then find the appropriate column based on the selected probability of .05 or .01, and a one- or two-tail test. Once we have found the appropriate column, the *t* critical value is found. If the calculated *t*-test value exceeds the *t* critical value from Appendix T, then the null hypothesis is rejected, and we conclude the sample mean difference was not due to chance occurrence.

6 Steps for Determining the Significance of the Independent Samples t-test

1) **Adopt a statistical significance level (usually .05)**
2) **Choose a one- or two-tailed test based on the null hypothesis**
3) **Calculate the degrees of freedom using *N - 2***

4) **Use *Appendix T* to find the *t* critical value.** Note that Appendix T lists *t* critical values as positive, but these values would also be applied to negative associations, and thus should be read as both positive and negative (+/-).

5) **Ask whether the *t*-test value calculated for the study exceeds the *t* critical value.** If the study *t* value exceeds the *t* critical value – be it either negative or positive – then the finding is statistically significant and we reject the null hypothesis. The resulting difference between the sample means exceeds what would be expected based on the adopted probability level. If the study *t* value does not exceed the *t* critical value, then the mean difference is not statistically significant and the null hypothesis is accepted.

6) **Interpret the results.** In addition to reporting whether the resulting *t*-test value is significant, interpreting results should also include the direction of the finding or confirmation of the null hypothesis. For example, stating there is a significant different between men and women and the sexual jealousy they report is only a partial interpretation. A more complete interpretation would include which group reported more sexual jealousy.

An alternative way to assess the significance of the independent samples *t*-test is to use a statistical program such as IBM SPSS or SAS. The statistical programs provide the exact probability of getting the calculated *t*-test value given the null hypothesis is true.

Effect Strength

Once the independent samples *t*-test is derived and assessed, the *t*-test may be converted to *r* as a measure of effect strength. Using Formula 9.3 presented earlier in the one-sample *t*-test section provides an effect size *r* to evaluate the impact or effect of group membership on the resulting mean difference on the dependent variable or outcome variable.

Applying the Independent Samples t-test to Case Study 9.3: Gender Differences in Sexual Jealousy

To help illustrate how the independent sample *t*-test works, we apply the *t*-test formula to our male and female sexual jealousy comparison. The null hypothesis is that male and female sexual jealousy means will be equal, $H_0: \mu_1 = \mu_2$.

Numerator

For the numerator of the *t*-test, the mean for Males (M = 3.6) is subtracted from the mean for Females (M = 2.8). This produces a mean difference of 0.80: 3.6 – 2.8 = 0.80.

Denominator

For the denominator, we first need to calculate the pooled variance (s_p^2). Following Formula 9.5, we take n_1 -1, which is 15-1 = 14, and multiply that by the variance for Males (s_1^2 = 0.83). This yields a value of 11.62: 14 x 0.83 = 11.62. We do the same for Females, taking n_2 -1 which is 15-1 = 14, and multiply that by the variance for Females (s_2^2 = 1.17). This yields a value of 16.38: 14 x 1.17 = 16.38. Next, add the values of 11.62 and 16.38, which gives us a numerator for the s_p^2 formula of 28: 11.62 + 16.38 = 28.0. The value of 28.0 is divided by n_1 + n_2 - 2 (which is 15 + 15 – 2 = 28), producing a pooled variance (s_p^2) of 1.0: 28/28 = 1.0. The calculations are shown below:

Steps *Pooled Variance*

1) $$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

2) $s_p^2 = \frac{(15-1)0.83+(15-1)1.17}{15+15-2}$

3) $s_p^2 = \frac{(14)0.83+(14)1.17}{28}$

4) $s_p^2 = \frac{(11.62)+(16.38)}{28}$

5) $s_p^2 = \frac{28}{28} = 1.0$

Now that we have the pooled variance s_p^2 of 1.0, we can calculate the denominator of the *t*-test. Take the pooled variance of 1.0 and divide by the number of cases for Males, which results in a value of .0667: 1.0/15 = .0667. The pooled variance of 1.0 is again divided by the number of cases for Females. This also gives us a value of .0667: 1.0/15 = .0667. Next, take the square root of the sum of these terms. This gives us a value of 0.365: $\sqrt{.0667+.0667} = 0.365$. Therefore, the denominator for the *t*-test formula is 0.365. The calculations are presented below:

Steps Denominator Independent Samples t-test formula

1) $\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}$

2) $\sqrt{\frac{1.0}{15}+\frac{1.0}{15}}$

3) $\sqrt{.0667+.0667}$

4) $\sqrt{0.1334} = 0.365$

Now take the mean difference from the numerator (0.80) and divide by the standard error of the difference between the means (0.365), which yields the final independent samples *t*-test value of 2.19: 0.80/0.365 = 2.19. The calculations are presented below:

Steps Independent Samples t-test formula

1) $$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

2) $$t = \frac{3.6 - 2.8}{0.365}$$

3) $$t = \frac{0.80}{0.365} = 2.19$$

The independent samples *t*-test value is 2.19, which may now be assessed for statistical significance.

Assessing the Statistical Significance for Case Study 9.2

Steps for Finding the t-test for Independent Samples Critical Value

1) Adopt a statistical significance level (usually .05)

- For the current example, we will adopt the .05 level of significance, indicating that we are willing to accept a less than 5% chance that our test results could have occurred by chance

2) Choose a one- or two-tail test based on the null hypothesis

- A two-tail test is chosen based on the null hypothesis

3) Calculate the degrees of freedom using *N - 2*

- For our current example, the degrees of freedom (N – 2) is 28: 30 – 2 = 28

4) Use *Appendix T* to find the *t* critical value

- Using a two-tailed test with a .05 level of significance, Appendix T shows a critical *t*-test value of +/- 2.05.

5) Ask whether the *t*-test value calculated for the study exceeds the *t* critical value

-Does our *t*-test value of 2.19 exceed the *t* critical value of +/- 2.05? It does, and we may conclude that our mean difference was not due to chance occurrence.

6) Interpret the results

-Based on the steps above, we conclude that Males differ from Females in regard to sexual jealousy. Further, we want to assess which group mean is greater. In this example, Males have a higher mean jealousy value than females. We reject the null hypothesis that the groups are equal.

A second way to make our statistical evaluation is to analyze the data using IBM SPSS or SAS. The applied examples section at the end of the chapter illustrates the IBM SPSS and SAS computer output for the current example. IBM SPSS shows the probability of getting a *t*-test value of 2.19 is .037, while SAS is more exact (p = .0369). Thus we reject the null hypothesis.

Effect Strength

Using Formula 9.3, an effect size is produced which indicates how big an effect gender has on the resulting sexual jealousy scores.

Step	Formula
1)	$r = \sqrt{\frac{t^2}{t^2 + df}}$
2)	$r = \sqrt{\frac{2.19^2}{2.19^2 + 28}}$
3)	$r = \sqrt{\frac{4.796}{4.796 + 28}}$
4)	$r = 0.38$

The resulting effect size based on r is 0.38, a medium effect. Gender has a medium effect on the resulting difference in sexual jealousy scores.

Example Write-up

A study was conducted to evaluate whether males and females differ in sexual jealousy. Based on prior research, the research hypothesis is that males would report higher levels of sexual jealousy than females. The null hypothesis is that sexual jealousy in males and females will be the same, H_0*:* $\mu_1 = \mu_2$. There *were 30 individuals recruited from an Introductory Psychology course for the study; 15 males and 15 females. An independent samples t-test was conducted to compare the sexual jealousy of males and females. The resulting mean difference was found to be statistically significant beyond the .05 level of significance,* $t(28) = 2.19$, $p < .05$, *with a medium effect* ($r = .38$)*. Thus we reject the null hypothesis. Overall, males were found to report greater sexual jealousy than females (M = 3.6 vs. M = 2.8, respectively).*

Assumptions of the t-test

For the *t*-test, there are a number of assumptions that should be met regarding the underlying distribution that the data form.

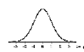
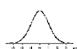
The first is *independence of observations*. This assumption refers to the relatedness of cases within each group. If cases have an underlying dependency, then the calculated *t*-test value may be inaccurate. For example, suppose you want to know if parents give more allowance to a male child than a female child. You might go to a mall and approach people who are shopping with a child and give them a brief questionnaire about allowances. You might be tempted to give the questionnaire to everyone including both parents of the child if both are shopping. The problem here is that the data within each group will not be independent – the information from the two parents is obviously related.

The problem can also arise in more subtle ways. A high school teacher might have one U.S. History class form study groups for a weekly in-class study session; another class has a weekly study session for individual students. Now the professor can compare exam scores of students in the two study conditions. A problem here is that the scores are not independent in the study group condition – the students in each of the study groups worked together and studied the same material. In this case, the data in the study group condition would be the average score of the groups. In the individual study class, students would be assigned to "pseudo-groups" with the data consisting of the average score of students in these groups.

For Case Study 9.3 regarding male and female sexual jealousy, can you think of some possible ways study participants might be interdependent? We can think of two (although you may have thought of others). One is that study participants might be dating one another and therefore would not be independent. Another would be that study participants might be siblings (brothers and sisters). During study recruitment, we want to screen study participants to make sure they were not dating anyone else in the study, nor should their siblings in the study.

A second assumption consists of three concerns -- *equal sample size, normality, and homogeneity of variance*. These three are typically covered as separate assumptions in many textbooks. However, because these three issues directly influence each other, we present them as a group. Generally, if you violate all three, then the *t*-test value you derive can be incorrect (i.e., biased or misleading). If only one or two of these concerns are violated, the *t*-test value should be accurate.

Equal sample size refers to having the same number of cases in each group. This is assessed by evaluating the number of cases in each group. If the same number of cases are in each group, then you have equal sample size. In practice, we think it is okay to have a few more cases in one group than another. For example, having group sizes of 52 and 50 is close enough, but having group sizes of 80 and 50 may be cause for concern.

Normality refers to the shape of the distribution of scores in the dependent variable for both groups. Normality concerns whether the distribution of scores <u>in each group</u> resembles a normal or bell-shaped distribution . Normality is assessed by plotting the data for each group. First, plot the data using a histogram and see if it forms a bell-shaped or normal distribution for each group . Next, use box plots and assess their shape. Both of these plots were covered in Chapter 3. Note that we recommend avoiding the use of statistical tests of normality such as the Kolmogorov-Sminov or Shapiro-Wilks tests found in IBM SPSS. Such tests can be overly sensitive to departures from normality. If you do use these tests, use a .001 level of significance as the probability cutoff.

Homogeneity of variance refers to the equality of the calculated variances for each group. If the homogeneity of variance is met, this means that the variances in each group are approximately the same. The best way to assess for homogeneity of variance is to compare the variances in each group. Generally, unequal variances for the *t*-test will not be problematic unless the variances are extremely different. For example, if one variance is more than 10 times that of the other variance, then the variances are said to be extremely different (Keppel, 1982). If the variance in one group is more than 10 times that in the other group, then the assumption of homogeneity of variance is violated. Further, according to some researchers, formal preliminary variance tests in most instances are not required (Markowski & Markowski, 1990). For example, statistical programs such as IBM SPSS will automatically produce one such test, called Levene's test. Such tests can be overly sensitive to variance differences. If you do use a test such as Levene's, use a .001 probability level to assess violation of the assumption.

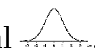
Let's reflect for a moment on how these issues are related. Recall that larger samples will produce distributions that are more normal. Therefore, a group that has a larger sample size will exhibit more normality. The greater the normality, the more likely a group mean will be representative of the scores. A small sample will be less likely to have a mean that is representative. And, variances for smaller sample sizes will be different from those taken from larger samples. Further, all three of these concerns are part of the *t*-test formula. The mean values are assumed to be from normally distributed samples and therefore are an accurate estimate of an individual's score. The standard error of the difference between the means is calculated using sample sizes and standard deviation values.

Although rare, violation of all three of these concerns does happen. The best approach when all three are violated is to use the *t*-test formula with the *separate variance estimation* calculation in the denominator. The *t*-test formula (9.7) using the separate variance estimate in the denominator is as follows:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Formula 9.7: Separate Variance Estimate t-test formula

Notice in this formula that the separate group variances are used instead of a pooled variance value? By using this approach, the different variances are taken into account, therefore providing a more accurate *t*-test value when all three of the assumptions (equal sample size, normality, equality of variances) are violated. In addition to a different denominator, degrees of freedom are calculated in a different fashion. If group sizes are exactly the same, this formula will produce the same *t*-test value as the pooled denominator *t*-test formula. Therefore, some researchers will use either formula when the group sizes are the same. IBM SPSS and SAS will report both the pooled and the separate variance estimate *t*-tests (the latter is reported with a label noting the variances are not equal).

For Case Study 9.3 addressing gender differences and sexual jealousy, the sample sizes were the same for each group, and the variances were well within the 10:1 ratio suggesting equality. In addition, the histograms for both groups (Figures 9.3 and 9.4) appear normal . Since at least one of these assumptions was met (in fact, all three were met), the resulting t-test is an accurate assessment of the differences between the sample means.

Note that the above assumptions apply to the independent sample t-test. For the one-sample t-test, only (a) independence of observations and (b) normality apply (because we have a single sample).

Case Study 9.4: Does a Seminar on the Benefits of Exercise Increase Exercise Behavior?

We present a second worked example illustrating results loosely based on a study by Dailey (2001). In her study, Dailey applied a health-behavior change model to increase exercise behavior in a group of young adults. The study was an experiment with two groups. The intervention group was given a 30-minute seminar on the benefits of exercise and how one can incorporate exercise into their daily schedule, and were asked to climb stairs instead of taking an elevator to underscore the importance of exercise behavior in everyday life. The control group was given a 30-minute seminar on how to improve study habits and received no information about exercise. Exercise behavior, measured in hours of exercise in a one-week period, was the outcome measure and was assessed 3-weeks after they attended the 30-minute seminar.

The research hypothesis (H_1) was that individuals who received the 30-minute seminar on the benefits of exercise would report higher levels of exercise behavior compared to the control group. The null hypothesis (H_0) is that the two groups would not differ in exercise behavior, $H_0: \mu_1 = \mu_2$, with μ_1 being the mean of the control group, and μ_2 the mean of the intervention group.

Data similar to Dailey's are presented in Table 9.5. The scores represent the number of hours exercised by each group summed across a 1-week period.

Summary Statistics: Means, Variances, and Standard Deviations

As with the previous example, the mean, variance, and standard deviation for both groups are presented. These summary statistics are noted in Table 9.6, and are discussed below.

Mean. For those in the control group, the mean number of hours exercising in a week is 10, while for those in the intervention group the mean value is 9. Note that contrary to our research hypothesis, the intervention group mean value is <u>lower</u> than the control group.

Control Group Mean: $M = \sum X_i/n = 200/20 = 10.0$

Intervention Group Mean: $M = \sum X_i/n = 180/20 = 9.0$

Variance. The variance for control group members is 27.05, and for the intervention group is 5.47.

Variance for Control Group: $$s^2 = \frac{\sum (X_i - M)^2}{n-1} = \frac{514}{19} = 27.05$$

Variance for Intervention Group: $$s^2 = \frac{\sum (X_i - M)^2}{n-1} = \frac{104}{19} = 5.47$$

Standard deviation. The standard deviations are noted below for each group, representing the average spread of scores for the control group (5.20) and for the intervention group (2.34).

Standard Deviation for Control Group: $$s = \sqrt{s^2} = \sqrt{27.05} = 5.20$$

Standard Deviation for Intervention Group: $s = \sqrt{s^2} = \sqrt{5.47} = 2.34$

Descriptive Data Plots

As we noted earlier, bar charts are an excellent way to visually exam the group mean values. Figure 9.5 is a bar chart, and presents the mean values for both the control group and the intervention group, with the control group mean being slightly higher. The histograms for each group (Figures 9.6 and 9.7) reveal normal distributions in hours of exercise per week.

Assumptions

Prior to assessing group differences in weekly exercise behavior with the *t*-test, we should formally assess the data assumptions. For independence of observations, we insure during data collection that cases are independent by utilizing a series of screening questions (e.g., "Is anyone who is related to you a participant in the current study?"). The sample sizes were the same for each group, and the variances (27.05 for the control group, and 5.47 for the intervention group) were well within the 10:1 ratio suggesting equality. In addition, the histograms for both groups (Figures 9.6 and 9.7) appear normal, suggesting the assumption of normality is met. A box plot is also generated (Figure 9.8) which also suggests normality (although there is one possible outlying case in the intervention group). Overall, we conclude that the data meet the assumptions of the independent samples t-test.

Applying the Independent Samples t-test to Case Study 9.4: Does a Seminar on the Benefits of Exercise Increase Exercise Behavior?

Now that basic descriptive statistics, plots, and the assumptions of the independent samples *t*-test have been evaluated, we move forward and formally calculate the t-test value.

Numerator

For the numerator of the *t*-test, we take the mean for those in the control group (M = 10.0) and subtract from the mean for those in the intervention group (M = 9.0). This produces a numerator mean difference of 1.0: 10.0 – 9.0 = 1.0.

Denominator

For the denominator, we first calculate the pooled variance (s_p^2). To do this, we will need the variances and sample sizes for each group:

Steps	*Pooled Variance for Independent Samples t-test*
1	$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
2	$s_p^2 = \frac{(20-1)27.05 + (20-1)5.74}{20+20-2}$
3	$s_p^2 = \frac{(19)27.05 + (19)5.74}{38}$
4	$s_p^2 = \frac{(513.95) + (109.06)}{38}$
5	$s_p^2 = \frac{623.01}{38}$ = 16.40 (rounded from 16.395)

Now that we have the pooled variance s_p^2 of 16.40, we calculate the denominator of the *t*-test statistic.

Steps	*Denominator t-test formula*
1	$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$
2	$\sqrt{\frac{16.4}{20} + \frac{16.4}{20}}$
3	$\sqrt{0.82 + 0.82}$

4 $\sqrt{1.64} = 1.28$

Next, take the mean difference from the numerator (1.0) and divide by the standard error of the difference between the means (1.28). This provides the final independent samples *t*-test value of 0.78: 1.0/1.28 = 0.78. The *t*-test statistic value is 0.78. The calculations are presented below:

Steps *t-test formula*

1 $$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

2 $$t = \frac{10.0 - 9.0}{1.28}$$

3 $$t = \frac{1.0}{1.28} = 0.78$$

Assessing the Statistical Significance of Case Study 9.4

Now that we know the *t*-test value is 0.78, we derive a *t* critical value for comparison. As in the prior examples, we walk through the 6 steps for finding a *t* critical value.

Steps for Finding the t-test for Independent Samples Critical Value

1) Adopt a statistical significance level

- We will adopt the .05 level of significance

2) Choose a one- or two-tail test based on the null hypothesis

- A two-tailed test is chosen

3) Calculate the degrees of freedom using *N - 2*

- The degrees of freedom for this example, based on *N – 2,* is 38: 40 – 2 = 38

4) Use Appendix T to find the *t* critical value

- Using a two-tailed test with a .05 level of significance, we use Appendix T to find the *t* critical value at 38 degrees of freedom. Notice that the table only has entries at 35 and at 40 degrees of freedom? When a critical value table does not include every possible degrees of freedom, the general rule is to pick the more conservative entry (in this instance, at 35 degrees of freedom). At 35 degrees of freedom, the *t* critical value is +/- 2.03, while at 40 degrees of freedom the cutoff is +/- 2.02. Hence, we will use the *t* critical value at 35 degrees of freedom of +/- 2.03 since it is more conservative.

5) Ask whether the *t*-test value exceeds the *t* critical value

-Does the *t*-test value of 0.78 exceed the *t* critical value of +/- 2.03? It does <u>not</u>. Our observed mean difference does not exceed our probability level.

6) Interpret the results

-The observed difference between the control and intervention group was not significant at the $p < .05$ level. The small difference we did observe (with the control group showing a slightly higher level of weekly exercise) is likely to have been due to chance. Thus we accept the null hypothesis.

The second way to make our evaluation is to analyze the data using IBM SPSS or SAS. Using these data, the programs tell us that the probability of getting a *t*-test value of 0.78 is equal to .438, which <u>does not</u> exceed our selected probability level of $p < .05$. We can further state that there is a 43.8% probability that the observed mean difference between the groups is due to chance occurrence.

Effect Strength

Using Formula 9.3, an effect size is produced which indicates how big an effect group membership has on weekly exercise behavior.

Step	Formula
1)	$r = \sqrt{\frac{t^2}{t^2 + df}}$
2)	$r = \sqrt{\frac{0.78^2}{0.78^2 + 38}}$
3)	$r = \sqrt{\frac{0.6084}{0.6084 + 38}}$
4)	$r = 0.126$

The resulting effect size based on *r* is 0.126, a small effect. Group membership has a small effect on resulting difference in weekly exercise behavior.

Example Write-up

An experiment was conducted to increase exercise behavior in young adults. Two experimental groups were utilized. One group, the intervention group, was given a 30-minute seminar on the benefits of exercise and asked to incorporate climbing stairs to emphasize the importance of exercise behavior in everyday life. A second group, the control group, was given a 30-minute seminar on how to improve study habits and received no information about exercise. The research hypothesis (H_1) was that individuals in the intervention group (who received the 30-minute seminar on the benefits of exercise) would report higher levels of exercise behavior compared to the control group. The null hypothesis (H_0) is that the two groups would not differ in exercise behavior, H_0: $\mu_1 = \mu_2$. Overall, 40 individuals were recruited for the study; 20 for the control group and 20 for the intervention group. Exercise behavior, measured in hours of *exercise in a one-week period, was the outcome measure, and was assessed 3-weeks after they attended the 30-minute seminar. An independent samples t-test was used to assess the group differences in weekly exercise behavior.*

Prior to assessment, data were evaluated for the assumptions of the t-test, including independence of observations, equal group sizes and variances, and normality. Overall, all assumptions were met. Independence of observations was met through screening items used during data collection. Both groups had equal sample sizes (20 in each group), and variances for each group (27.05 for the control group and 5.47 for the intervention group) were within the 10:1 ratio recommendation. Histograms and a box plot showed both variables to be normally distributed.

The results of the independent samples t-test showed that the control group mean (M = 10.0) did not differ significantly from the intervention group mean (M = 9.0) at the .05 level of significance, t(38) = 0.78, n.s., and evidenced only a small effect (r = .126). We therefore accept the null hypothesis, and conclude the seminar on exercise behavior had no effect on exercise behavior. The small observed difference between the control and intervention groups is likely to have been due to chance.

Interesting Web-based Links

http://glass.ed.asu.edu/stats/analysis/ttest.html {One-Sample *t*-test}

http://physics.ubishops.ca/phy101/lectures/Beaver/twoSampleTTest.html {Independent Sample *t*-test using means, standard deviations, and group samples sizes for input}

Applied Examples Using IBM SPSS and SAS

In both IBM SPSS and SAS, the one sample and independent samples *t*-test can be performed easily. In IBM SPSS, for a one sample *t*-test, use the One Sample *t*-test procedure. For the independent samples *t*-test, use the *Independent Samples t-test* procedure. Both are found in the pull-down menu under *Analyze* then *Compare Means*. Syntax may also be used. In SAS, use *Proc TTest* for both procedures with different subcommands.

One-Sample t-test

Here, we illustrate the one-sample *t*-test on *Case Study 9.1*; television viewing in a 24-hour period by college students compared to the average American adult. For IBM SPSS and SAS, the data would be entered looking like this (from Table 9.1).

Data from Table 9.1: Hours per day college students watch television (N = 10)

Student	Hours College Students Watch Television (24-Hour Period)
1	2
2	4
3	4
4	4
5	4
6	4
7	3
8	3
9	5
10	6

We have abbreviated the label for television viewing as *TVperday*, and will add a *variable label* to read *Television viewing per day*.

Example IBM SPSS syntax is presented below to compare this data to the average adult value of 5 hours per day of viewing television. Open a syntax window (go to *File*, then *New*, then *Syntax*). You may directly enter the *t*-test commands shown on the next page in the syntax window. Then, highlight all the syntax, and press the *Play* icon which looks like a play button on a DVD or VCR). Your output should resemble what is illustrated on the next page.

In syntax, we use a “Test Value” command to note the population mean of “5” that we will be comparing to our sample mean. Note that IBM SPSS provides descriptive statistics for the sample (mean, standard deviation) and the standard error which ultimately is the denominator in the one-sample *t*-test formula. The program also provides 95% confidence intervals for the mean difference.

```
T-TEST
  /TESTVAL=5
  /MISSING=ANALYSIS
  /VARIABLES=TVperday
  /CRITERIA=CI(.95).
```

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
TVperday	**10**	**3.9000**	**1.10050**	**.34801**

One-Sample Test

	Test Value = 5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
TVperday	**-3.161**	**9**	**.012**	**-1.10000**	**-1.8873**	**-.3127**

For SAS, we use Proc TTest, and use a command called "H0" in the syntax line. This is done to indicate the population mean to which we are comparing our sample mean. Here, $H_0 = 5$, indicating the 5 hours of television viewing for the average adult per day. SAS produces descriptive statistics including the sample mean and standard deviation, and the standard error. Both 95% confidence intervals for the sample mean and standard deviation is also provided in the output.

```
Data TVexample;
input Student TVperday;
Datalines;
1      2
2      4
3      4
4      4
5      4
6      4
7      3
8      3
9      5
10     6
;
run;

proc ttest H0=5;
var TVperday;
run;
```

The TTEST Procedure

Statistics

Variable	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Minimum	Maximum
TVperday	10	3.1127	3.9	4.6873	0.757	1.1005	2.0091	0.348	2	6

T-Tests

Variable	DF	t Value	Pr > |t|
TVperday	9	-3.16	0.0115

Independent Samples t-test

We illustrate the independent samples *t*-test using data from *Case Study 9.3*, comparing males and females on their reports of sexual jealousy. The original data is presented below (from Table 9.3).

Data from Table 9.3: Sexual Jealousy scores for Males and Females (N = 30)

	Group 1 (Males)	Group 2 (Females)
	3	3
	3	3
	2	3
	2	3
	3	2
	4	1
	4	1
	3	2
	4	2
	4	3
	4	3
	5	3
	5	5
	4	4
	4	4
N	15	15

This is the data set from the jealousy example from Table 9.3. However, the data must be reorganized before it is entered into the IBM SPSS *Data View* data window and SAS syntax edit screen. Both IBM SPSS and SAS expect that you will have a single column of data indicating your independent variable or grouping variable, and a single column of data indicating your dependent variable or outcome variable. The data should be reorganized into the following format:

Revised Data from Table 9.3: Sexual Jealousy scores by Males and Females (N = 30)

Group (Male = 1, Female = 2)	Jealousy
1	3
1	3
1	2
1	2
1	3
1	4
1	4
1	3
1	4
1	4
1	4
1	5
1	5
1	4
1	4
2	3
2	3
2	3
2	3
2	2
2	1
2	1
2	2
2	2
2	3
2	3
2	3
2	5
2	4
2	4

The above format has a single column indicating Group Membership. Males are coded as “1” while Females are coded as “2”. The other column of data indicates the sexual jealousy scores.

Once reorganized, the data may be entered directly into IBM SPSS or SAS. Only enter the numeric data. Once you have entered the two columns of data, in IBM SPSS you can go to the *Variable View* screen and enter the two variable names (Group, Jealousy) and fill-in the variable and value labels. For example, for the variable *Group* you can enter *Gender Group* as the variable label, and enter *1 = Male* and *2 = Female* as the value label. For the variable *Jealousy*, you can enter a variable label *Jealousy Score*.

IBM SPSS syntax. Once data are entered in IBM SPSS, we can now proceed with the analysis using programming syntax. Open a syntax window (go to *File*, then *New*, then *Syntax*). You may directly enter the *t*-test commands shown on the next page in the syntax window. Then, highlight all the syntax, and press the *Play* icon which looks like a play button on a DVD or VCR). Your output should resemble what is illustrated on the next page.

```
T-TEST GROUPS=Group(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=Jealousy
  /CRITERIA=CI(.95).
```

T-Test

Group Statistics

	Group Gender	N	Mean	Std. Deviation	Std. Error Mean
Jealousy Score	1 Male	**15**	**3.60**	**.910**	**.235**
	2 Female	**15**	**2.80**	**1.082**	**.279**

Independent Samples Test

		Levene's Test for Equality of Variances		*t*-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Jealousy Score	Equal variances assumed	**.060**	**.809**	**2.191**	**28**	**.037**	**.800**	**.365**	**.052**	**1.548**
	Equal variances not assumed			**2.191**	**27.201**	**.037**	**.800**	**.365**	**.051**	**1.549**

IBM SPSS first provides the group sample sizes, means and standard deviation values, and standard errors. Levene's test is next offered as an assessment of equality of variances between groups (if used, be conservative with this test at the $p < .001$ level). The *t*-test value is next provided. Next to "Equal variances assumed" is the pooled variance result, showing t(28) = 2.191, $p = .037$ (two-tailed). The mean difference (0.80) is provided, as is a confidence interval surrounding the mean difference.

SAS syntax. We use the exact same data setup noted in Table 9.3. As with most of our SAS examples, we provide the full data setup and output. Once you open SAS, simply go to your *Program Editor* window, and type the following syntax:

```
Data example;
input Group Jealousy;
Datalines;
1      3
1      3
1      2
1      2
1      3
1      4
1      4
1      3
1      4
1      4
1      4
1      5
1      5
1      4
1      4
2      3
2      3
2      3
2      3
2      2
2      1
2      1
2      2
2      2
2      3
2      3
2      3
2      5
2      4
2      4
;
run;

Proc ttest;
  class Group;
  var Jealousy;
run;
```

The SAS System

The TTEST Procedure

Statistics

Variable	Group	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Minimum	Maximum
Jealousy	1	15	3.0959	3.6	4.1041	0.6664	0.9103	1.4356	0.235	2	5
Jealousy	2	15	2.2006	2.8	3.3994	0.7924	1.0823	1.7069	0.2795	1	5
Jealousy	Diff (1-2)		0.052	0.8	1.548	0.7936	1	1.3525	0.3651		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
Jealousy	Pooled	Equal	28	2.19	0.0369
Jealousy	Satterthwaite	Unequal	27.2	2.19	0.0372

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
Jealousy	Folded F	14	14	1.41	0.5255

SAS provides the group sample sizes, means and standard deviation values (and 95% confidence intervals), and standard errors.

Equality of variances is assessed using a Folded F method. Our *t*-test value is located in the center of the output, next to *Pooled*.

Below that is the unequal variance *t*-test, labeled *Satterthwaite* to indicate the method of adjustment for unequal variances.

Draft Sample Exercises (additional exercises will be created)

Problem Set 1: An investigation was performed comparing the ages of individuals who watch the television show “American Idol” to those who do not watch the show. Are those who watch “American Idol” younger than those who do not watch? Perform an independent samples *t*-test on the data to investigate. The data collected are represented below.

	Group 1 (Watch American Idol)	Group 2 (Do not Watch)
	18	21
	18	22
	18	20
	22	18
	21	19
	19	18
	21	18
	18	22
	23	20
N	9	9

A. Derive the null and research hypotheses for this problem.

B. Calculate the means, standard deviations, and variances for each group.

C. Graph the group means appropriately. Based on the graphed means, is one group younger than the other?

D. Assess the assumptions of the *t*-test (independence of observations, equal sample size, normality, equality of variance). Are the assumptions met? How did you make this assessment?

E. Calculate the independent samples *t*-test assuming a two-tailed test at the .05 level of significance. What value did you calculate? Was it significant at the .05 level?

F. Write a brief summary of the project, including an overview of the data, hypothesis, and *t*-test findings (including mean values). Based on these findings, are those who watch “American Idol” significantly different in age from those who do not watch the show?

<u>Problem Set 2</u>: An experiment was conducted to investigate whether individuals' psychological well-being can be influenced by positive comments on a problem-solving task. The hypothesis is that positive comments would lead to greater well-being (measured on a well-being scale). Twenty individuals participated in the study, with 10 randomly assigned to each group. For each group, study participants were provided a series of anagrams to solve for 20 minutes. Those in the control group were provided no feedback on their work, while those in the intervention group were provided positive comments regardless of whether they completed the anagrams. At the end of the 20 minutes, participants in both groups were given a well-being measure, with higher scores indicating greater psychological well-being. Perform an independent samples *t*-test on the data provided (below):

	Group 1 Control	Group 2 Intervention
	21	19
	21	20
	20	21
	22	19
	20	18
	20	19
	19	17
	20	20
	19	17
	21	20
N	10	10

A. Derive the null and research hypotheses for this problem.

B. Calculate the means, standard deviations, and variances for each group.

C. Graph the group means appropriately. Based on the graphed means, which group has greater well-being?

D. Assess the assumptions of the *t*-test (independence of observations, equal sample size, normality, equality of variance). Are the assumptions met? How did you assess these assumptions?

E. Calculate the independent samples *t*-test, using a two-tailed test at the .05 level of significance. What value did you calculate? Is it significant at the .05 level?

F. Write a brief summary of the project, including an overview of the data, hypothesis, and *t*-test findings (including mean values). Do the two groups differ in well-being?

G. As a budding psychologist, think for a moment about the findings – what might explain the group differences?

Problem 3: In an effort to evaluate whether a new depression inventory can be used with various subject populations, a study was conducted to compare the mean value on the inventory from a group of younger adults to a "normative" mean established with older adults. The hope is that the younger adult mean value will not differ from the "normed" value for older adults (the normative mean value for older adults is 10.1 on the scale, with higher scores indicating greater depression, and a value of "13" indicative of clinical depression). Study participants were 11 younger adults. Each completed the new depression inventory. Perform a one-sample *t*-test on the data provided (below):

	Younger Adult Sample
	10.5
	10.2
	11
	13
	7
	10
	10
	11
	8
	10.5
	14
N	11

A. Derive the null and research hypotheses for this problem.

B. Calculate the mean, standard deviation, and variance for the group of young adults, and create a box-plot of the data. Based simply on this data, does the sample appear to match the normative mean value? By the way, are there any individuals in this sample that are clinically depressed? If so, who?

C. Evaluate a limited set of assumptions with this data (independence of observations, and normality). Are the assumptions met? How did you assess these assumptions?

D. Calculate the one-sample *t*-test using a two-tailed test at the .05 level of significance. What value did you calculate? Is it significant at the .05 level?

E. Write a brief summary of the project, including an overview of the data, hypothesis, and *t*-test findings (including your mean value). Is the mean from the young adult sample different from the normative mean from the older adults?

Appendix T

(<u>Draft</u> -- NOTE that one-tail table values will added at a later date)

	Two Tailed Significance							
Df	**0.2**	**0.1**	**0.05**	**0.01**	**0.005**	**0.001**	**0.0005**	**0.0001**
2	1.89	2.92	4.30	9.92	14.09	31.60	44.70	100.14
3	1.64	2.35	3.18	5.84	7.45	12.92	16.33	28.01
4	1.53	2.13	2.78	4.60	5.60	8.61	10.31	15.53
5	1.48	2.02	2.57	4.03	4.77	6.87	7.98	11.18
6	1.44	1.94	2.45	3.71	4.32	5.96	6.79	9.08
7	1.41	1.89	2.36	3.50	4.03	5.41	6.08	7.89
8	1.40	1.86	2.31	3.36	3.83	5.04	5.62	7.12
9	1.38	1.83	2.26	3.25	3.69	4.78	5.29	6.59
10	1.37	1.81	2.23	3.17	3.58	4.59	5.05	6.21
11	1.36	1.80	2.20	3.11	3.50	4.44	4.86	5.92
12	1.36	1.78	2.18	3.05	3.43	4.32	4.72	5.70
13	1.35	1.77	2.16	3.01	3.37	4.22	4.60	5.51
14	1.35	1.76	2.14	2.98	3.33	4.14	4.50	5.36
15	1.34	1.75	2.13	2.95	3.29	4.07	4.42	5.24
16	1.34	1.75	2.12	2.92	3.25	4.01	4.35	5.13
17	1.33	1.74	2.11	2.90	3.22	3.97	4.29	5.04
18	1.33	1.73	2.10	2.88	3.20	3.92	4.23	4.97
19	1.33	1.73	2.09	2.86	3.17	3.88	4.19	4.90
20	1.33	1.72	2.09	2.85	3.15	3.85	4.15	4.84
21	1.32	1.72	2.08	2.83	3.14	3.82	4.11	4.78
22	1.32	1.72	2.07	2.82	3.12	3.79	4.08	4.74
23	1.32	1.71	2.07	2.81	3.10	3.77	4.05	4.69
24	1.32	1.71	2.06	2.80	3.09	3.75	4.02	4.65
25	1.32	1.71	2.06	2.79	3.08	3.73	4.00	4.62
26	1.31	1.71	2.06	2.78	3.07	3.71	3.97	4.59

27	1.31	1.70	2.05	2.77	3.06	3.69	3.95	4.56
28	1.31	1.70	2.05	2.76	3.05	3.67	3.93	4.53
29	1.31	1.70	2.05	2.76	3.04	3.66	3.92	4.51
30	1.31	1.70	2.04	2.75	3.03	3.65	3.90	4.48
35	1.31	1.69	2.03	2.72	3.00	3.59	3.84	4.39
40	1.30	1.68	2.02	2.70	2.97	3.55	3.79	4.32
45	1.30	1.68	2.01	2.69	2.95	3.52	3.75	4.27
50	1.30	1.68	2.01	2.68	2.94	3.50	3.72	4.23
55	1.30	1.67	2.00	2.67	2.92	3.48	3.70	4.20
60	1.30	1.67	2.00	2.66	2.91	3.46	3.68	4.17
65	1.29	1.67	2.00	2.65	2.91	3.45	3.66	4.15
70	1.29	1.67	1.99	2.65	2.90	3.43	3.65	4.13
75	1.29	1.67	1.99	2.64	2.89	3.42	3.64	4.11
80	1.29	1.66	1.99	2.64	2.89	3.42	3.63	4.10
85	1.29	1.66	1.99	2.63	2.88	3.41	3.62	4.08
90	1.29	1.66	1.99	2.63	2.88	3.40	3.61	4.07
95	1.29	1.66	1.99	2.63	2.87	3.40	3.60	4.06
100	1.29	1.66	1.98	2.63	2.87	3.39	3.60	4.05
200	1.29	1.65	1.97	2.60	2.84	3.34	3.54	3.97
500	1.28	1.65	1.96	2.59	2.82	3.31	3.50	3.92
1000	1.28	1.65	1.96	2.58	2.81	3.30	3.49	3.91
Infinity	1.28	1.64	1.96	2.58	2.81	3.29	3.48	3.89

Table 9.1: Hours per day college students watch television (N = 10)

Student	Hours College Students Watch Television (24-Hour Period)
1	2
2	4
3	4
4	4
5	4
6	4
7	3
8	3
9	5
10	6

Table 9.2: Hours college students watch television (24-hour period), summary statistics, and one-sample t-test components (N = 10)

Student	Hours College Students Watch Television (24-Hour Period)	X - M	$(X - M)^2$
1	2	-1.90	3.61
2	4	0.10	0.01
3	4	0.10	0.01
4	4	0.10	0.01
5	4	0.10	0.01
6	4	0.10	0.01
7	3	-0.90	0.81
8	3	-0.90	0.81
9	5	1.10	1.21
10	6	2.10	4.41
Sum	39		10.90
Mean	3.90		
Variance	1.21		
SD	1.10		

Table 9.3: Sexual Jealousy scores for Males and Females (N = 30)

	Group 1 Males	Group 2 Females
	3	3
	3	3
	2	3
	2	3
	3	2
	4	1
	4	1
	3	2
	4	2
	4	3
	4	3
	5	3
	5	5
	4	4
	4	4
N	15	15

Table 9.4: Sexual Jealousy scores for Males and Females, summary statistics, and t-test calculation components (N = 30)

	Group 1 (Males)	X - *M*	(X - *M*)^2	**Group 2 (Females)**	X - *M*	(X - *M*)^2
	3	-0.60	0.36	**3**	0.20	0.04
	3	-0.60	0.36	**3**	0.20	0.04
	2	-1.60	2.56	**3**	0.20	0.04
	2	-1.60	2.56	**3**	0.20	0.04
	3	-0.60	0.36	**2**	-0.80	0.64
	4	0.40	0.16	**1**	-1.80	3.24
	4	0.40	0.16	**1**	-1.80	3.24
	3	-0.60	0.36	**2**	-0.80	0.64
	4	0.40	0.16	**2**	-0.80	0.64
	4	0.40	0.16	**3**	0.20	0.04
	4	0.40	0.16	**3**	0.20	0.04
	5	1.40	1.96	**3**	0.20	0.04
	5	1.40	1.96	**5**	2.20	4.84
	4	0.40	0.16	**4**	1.20	1.44
	4	0.40	0.16	**4**	1.20	1.44
Sum	54		11.60	42		16.40
Mean	3.60			2.80		
Variance	0.83			1.17		
SD	0.91			1.08		

Table 9.5: Weekly rate of exercise (in hours) for control and intervention group (N = 40)

	Group 1 (Control)	**Group 2 (Intervention)**
	6	9
	9	9
	10	9
	10	10
	10	10
	6	10
	2	9
	3	9
	4	4
	4	5
	6	6
	13	6
	15	8
	18	8
	18	13
	10	12
	10	12
	13	11
	13	10
	20	10
N	20	20

Table 9.6: Weekly rate of exercise (in hours) for control and intervention group, summary statistics, and t-test calculation components (N = 40)

	Group 1 (Control)	X - M	$(X - M)^2$	**Group 2 (Intervention)**	X - M	$(X - M)^2$
	6	-4.00	16.00	9	0.00	0.000
	9	-1.00	1.00	9	0.00	0.000
	10	0.00	0.00	9	0.00	0.000
	10	0.00	0.00	10	1.00	1.000
	10	0.00	0.00	10	1.00	1.000
	6	-4.00	16.00	10	1.00	1.000
	2	-8.00	64.00	9	0.00	0.000
	3	-7.00	49.00	9	0.00	0.000
	4	-6.00	36.00	4	-5.00	25.000
	4	-6.00	36.00	5	-4.00	16.000
	6	-4.00	16.00	6	-3.00	9.000
	13	3.00	9.00	6	-3.00	9.000
	15	5.00	25.00	8	-1.00	1.000
	18	8.00	64.00	8	-1.00	1.000
	18	8.00	64.00	13	4.00	16.000
	10	0.00	0.00	12	3.00	9.000
	10	0.00	0.00	12	3.00	9.000
	13	3.00	9.00	11	2.00	4.000
	13	3.00	9.00	10	1.00	1.000
	20	10.00	100.00	10	1.00	1.000
Sum	200		514	180		104
Mean	10.0			9.0		
Variance	27.05			5.47		
SD	5.20			2.34		

Figure 9.1: Histogram of hours per day college students watch television (N =10)

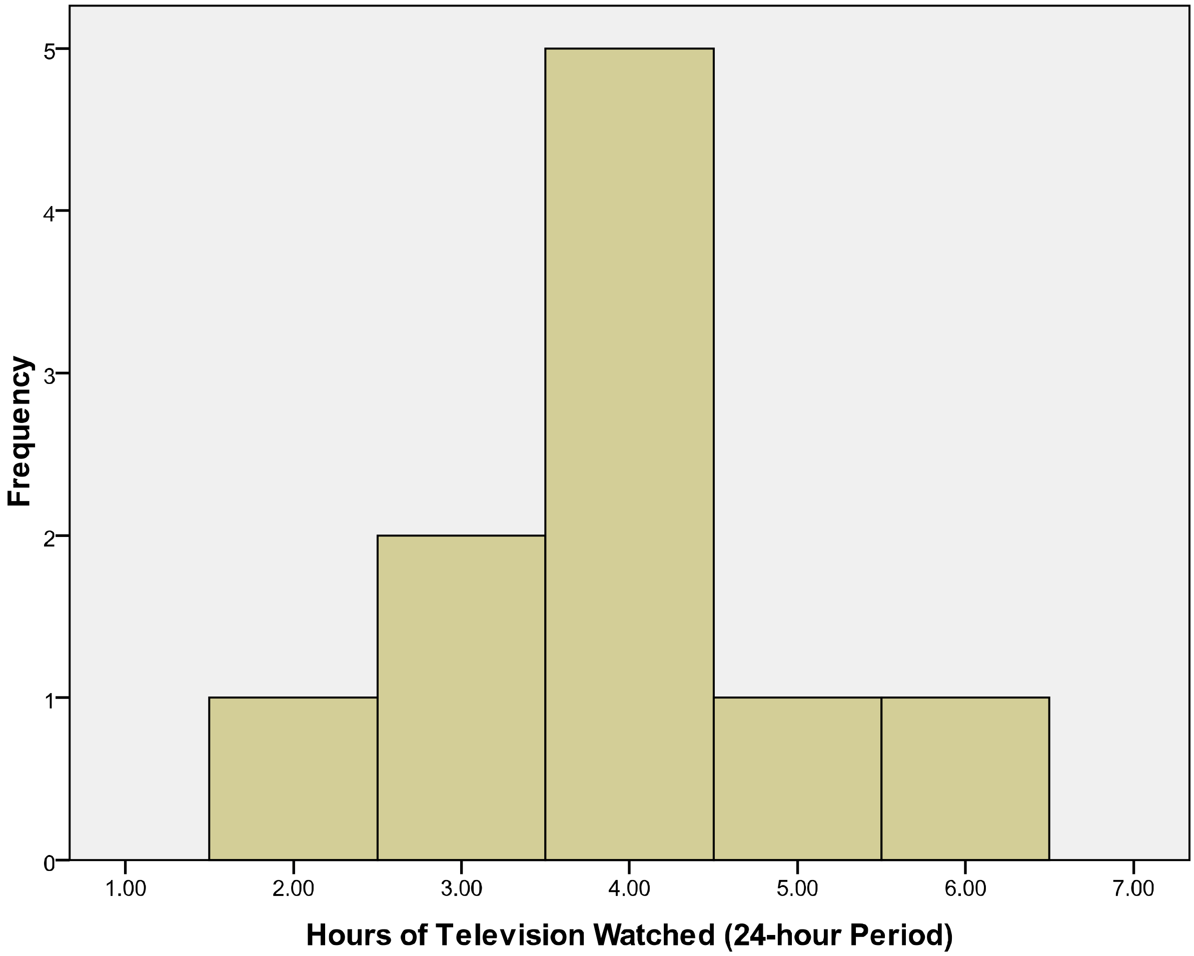


Figure 9.2: Bar chart of Mean Sexual Jealousy Scores by Gender

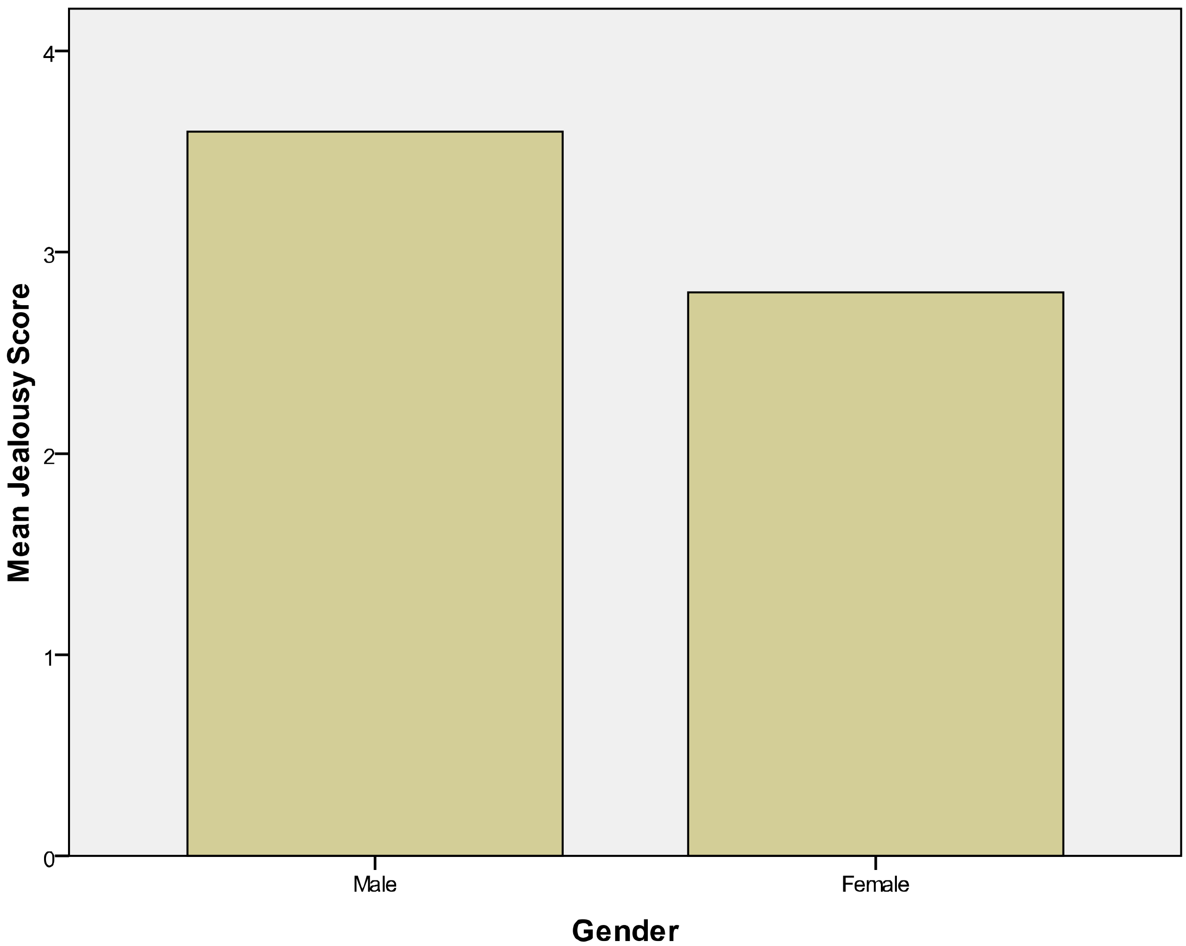


Figure 9.3: Histogram of Male Sexual Jealousy Scores

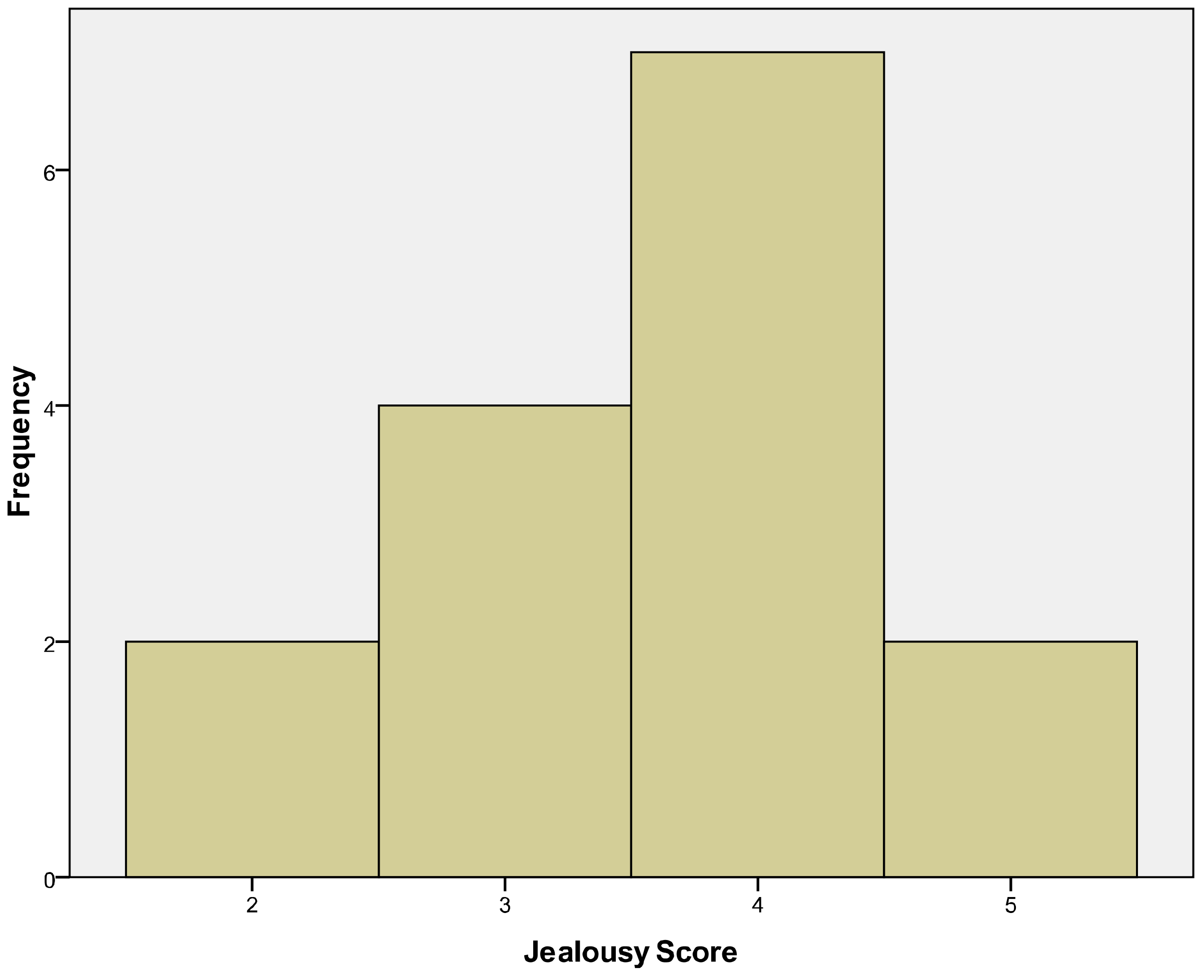


Figure 9.4: Histogram of Female Sexual Jealousy Scores

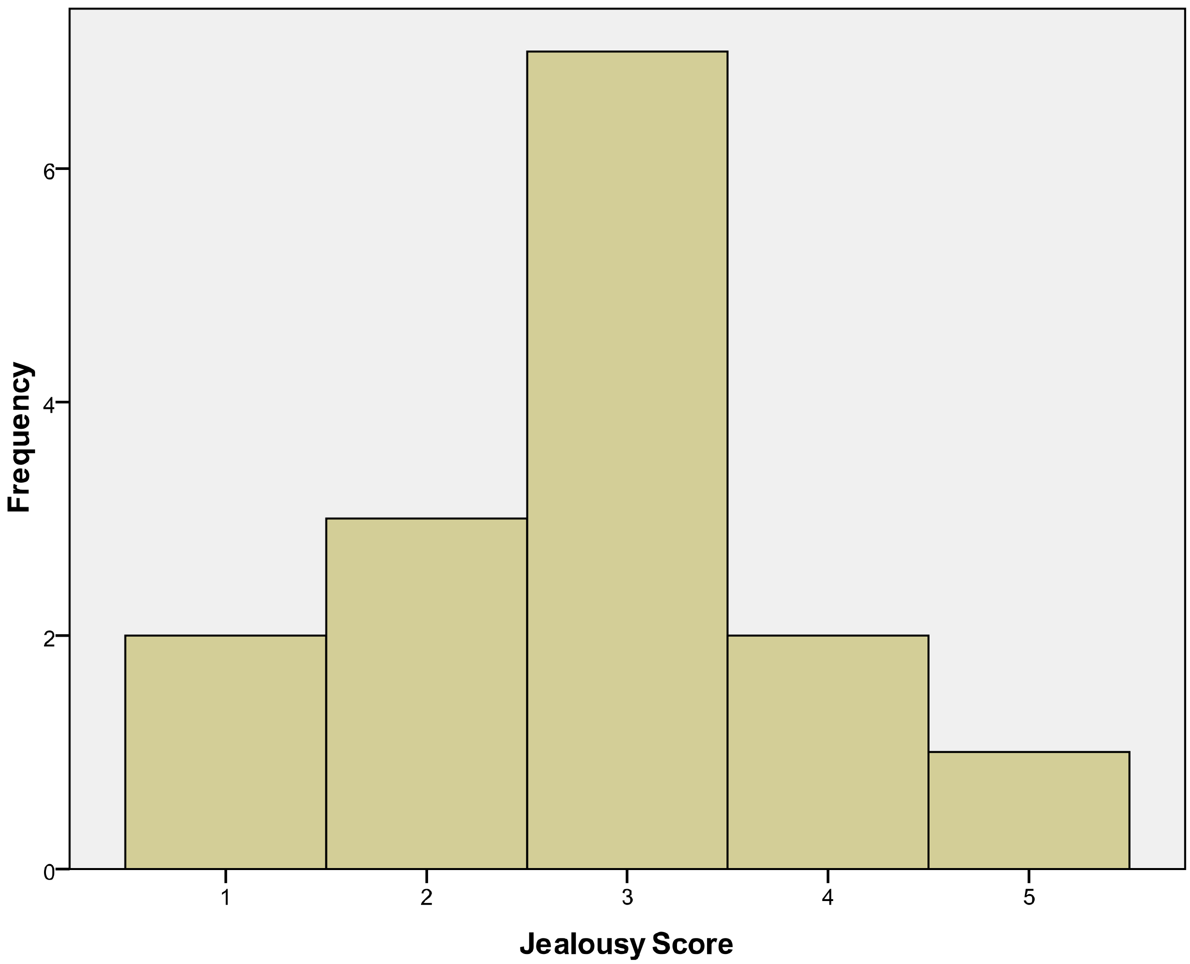


Figure 9.5: Bar chart of mean weekly rate of exercise (in hours) for control and intervention group

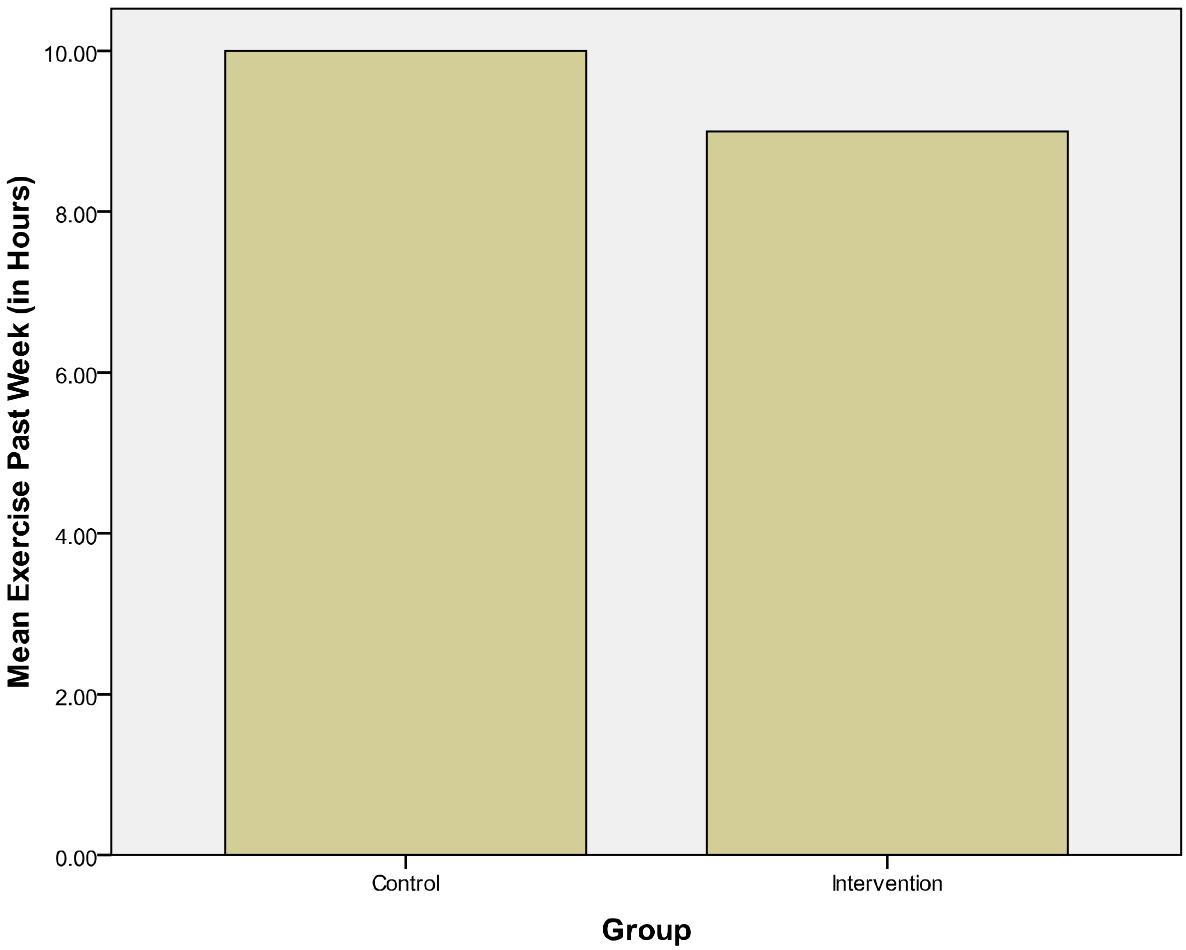


Figure 9.6: Histogram of weekly rate of exercise (in hours) for control group

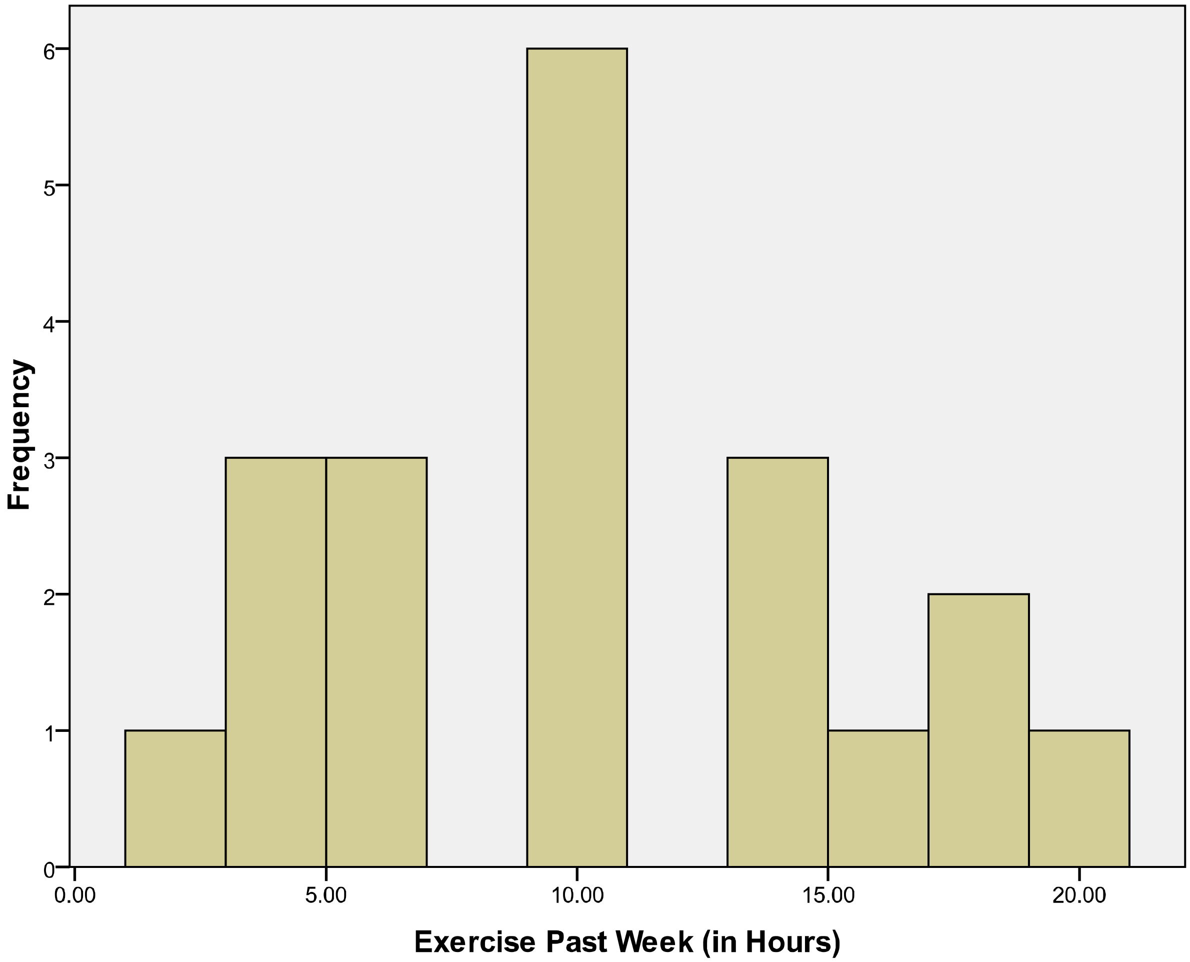


Figure 9.7: Histogram of weekly rate of exercise (in hours) for intervention group

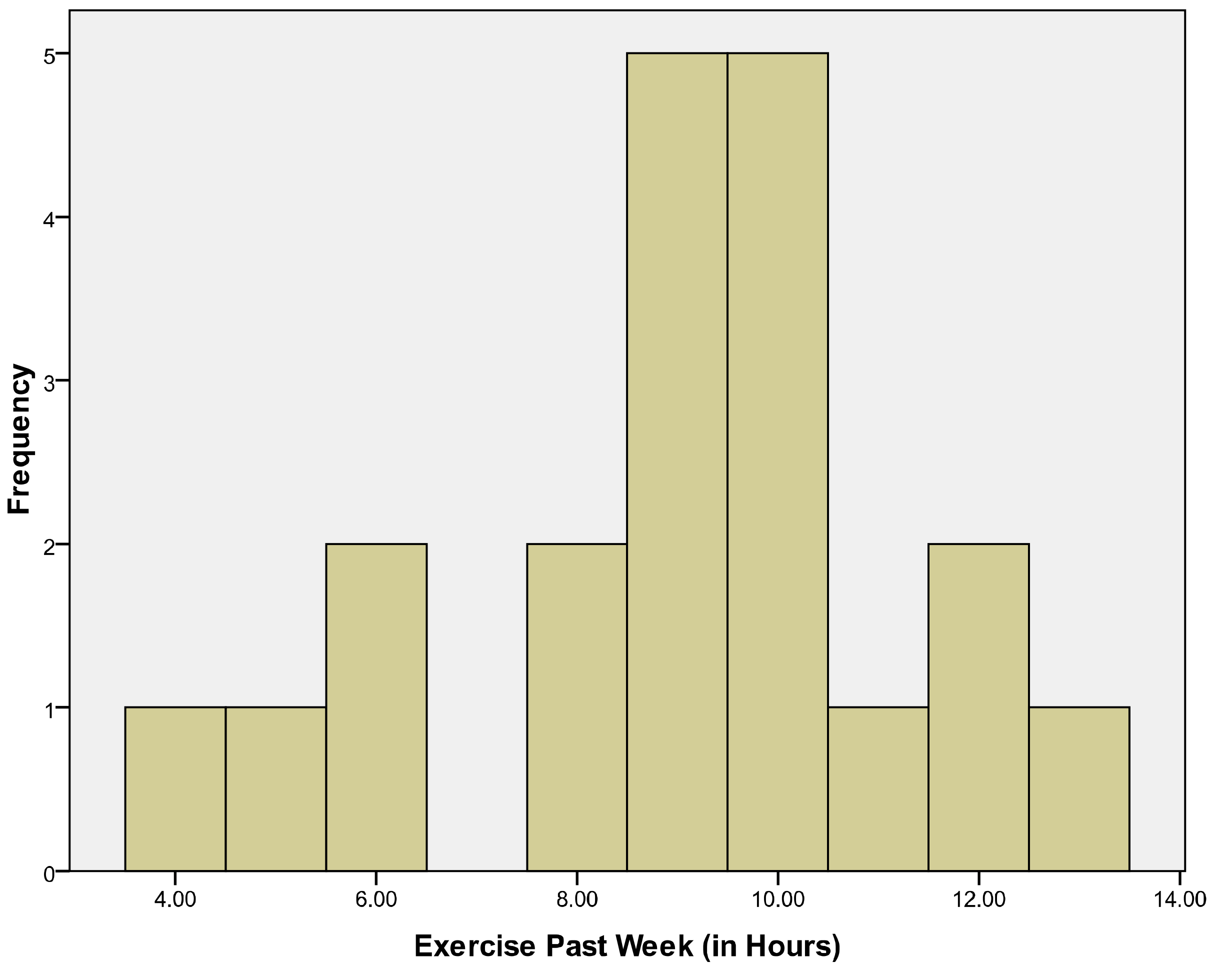


Figure 9.8: Box plot of weekly rate of exercise (in hours) for control and intervention groups