

**COMPARING TWO POPULATION MEANS: INDEPENDENT SAMPLES**

In Chapter 5, we considered inference for a single population mean. Then, in the first part of Chapter 6 we extended these ideas so that comparisons could be made between two groups. Such comparisons were made using differences because the observations in the two groups were related, or dependent.

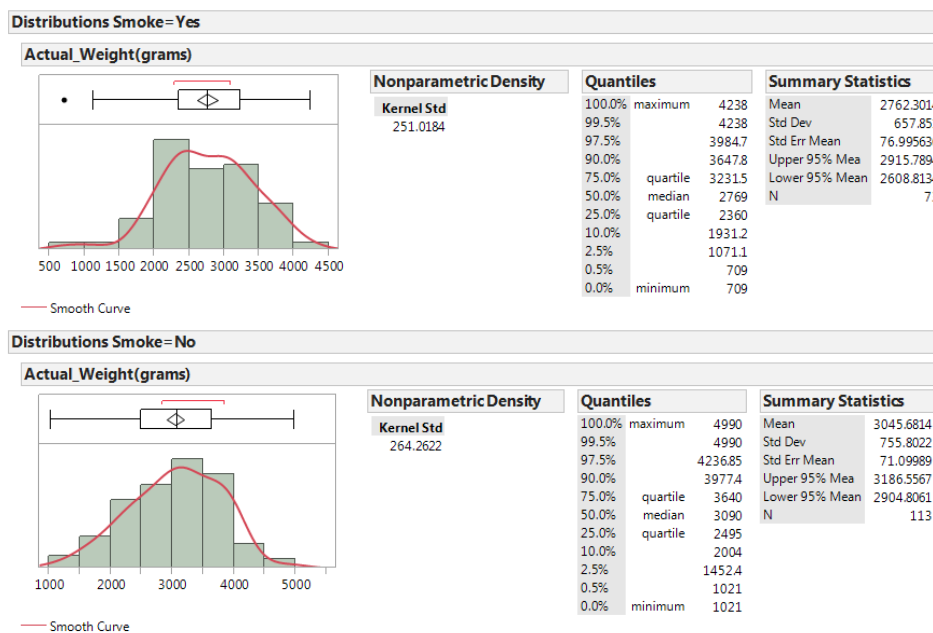
In this section, we will consider making comparisons between two groups when the observations in one group are independent of the observations in the other group. The methodologies considered here are a bit different because it no longer makes sense to simply work with differences. Consider the following example.

**Example 6.5:** A study was conducted in which the relationship between smoking and birth weights of babies was considered. The data can be found in the file **LowBirth.JMP**, a portion of which is shown below.

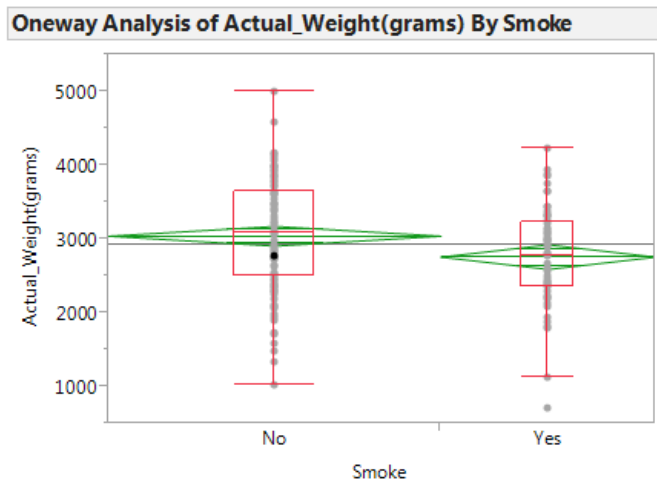
	Race	Previous_History	Hypertension	Smoke	Uterine_Irritation	Mothers_Age	Actual_Weight(grams)	Weight
1	Nonwhite	No	Normal	No	Yes	19	2523	Normal
2	Nonwhite	No	Normal	No	No	33	2551	Normal
3	White	No	Normal	Yes	No	20	2557	Normal

**Comment:** The first observation for a nonsmoker is in no way related to the first observation for a smoker. Therefore, these groups are independent.

Next, let's examine birth weight across the two groups: smokers and nonsmokers.



In JMP, we can compare the distributions by selecting **Analyze > Fit Y by X**. Place the categorical variable (**Smoke**) in the **X, Factor** box, and place the numerical variable of interest (**Actual\_Weight**) in the **Y, Response** box. Select the “Display Options” menu from the red drop-down arrow to display the boxplots and mean diamonds.



Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err		
				Mean	Lower 95%	Upper 95%
No	113	3045.68	755.802	71.100	2904.8	3186.6
Yes	73	2762.30	657.851	76.996	2608.8	2915.8

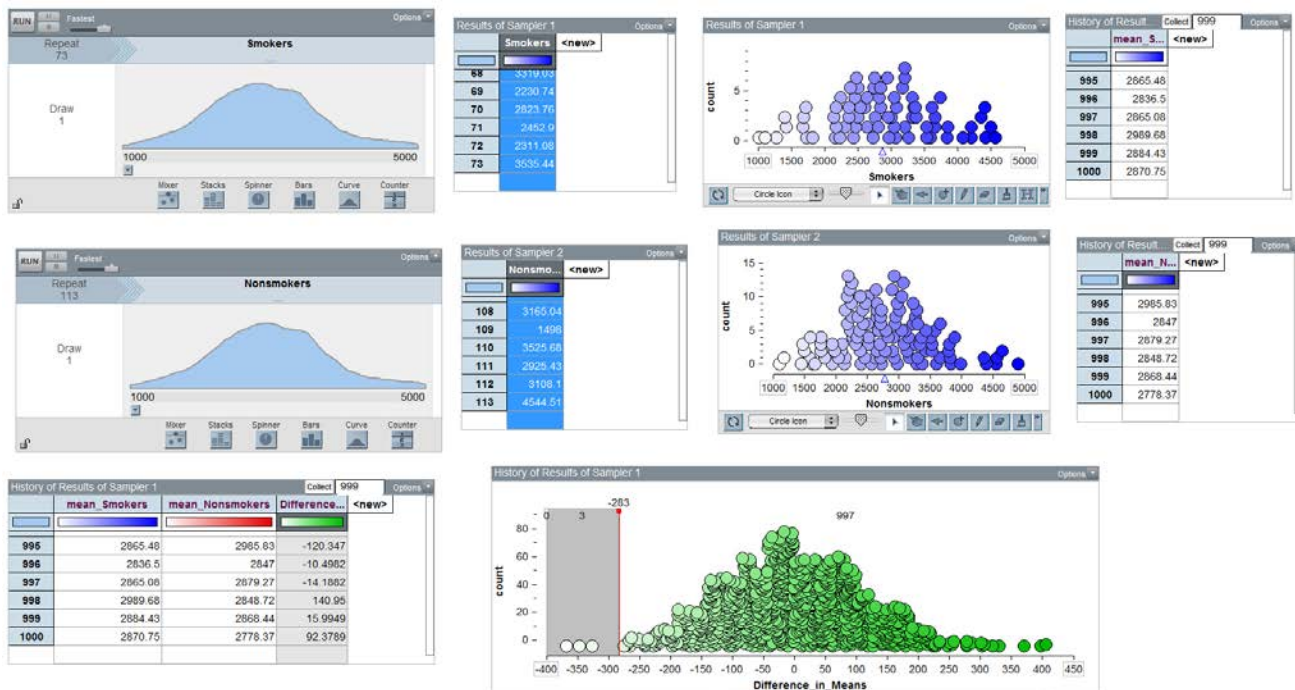
The sample statistics are summarized in the following table:

	Sample Statistics	
	Group 1 (Smokers = “Yes”)	Group 2 (Nonsmokers = “No”)
Mean	$\bar{x}_1 = 2762.30 \text{ g}$	$\bar{x}_2 = 3045.68 \text{ g}$
Standard Deviation	$s_1 = 657.85 \text{ g}$	$s_2 = 755.80 \text{ g}$
# of Observations	$n_1 = 73$	$n_2 = 113$

Questions:

1. We are interested in the true difference between these two population means,  $\mu_1 - \mu_2$ . Based on the data, what is your best guess for this difference, and what does it indicate about birth weights for smokers compared to nonsmokers?
2. If you were to take another sample, is your “best guess” from the previous question likely to change? Explain.

Since the difference will change from sample to sample, in order to make valid inferences about the true population difference, we must first understand *how* the difference in sample means is expected to change from sample to sample. That is, we must determine what differences in means are likely to happen by chance when taking random samples from populations with the same mean. The following illustration in Tinkerplots gives us some insight into this.



### Characteristics of the Sampling Distribution for the Difference in Means

The illustration in Tinkerplots helps us understand the intuition behind inference for a difference in means, but we don't really need Tinkerplots to determine the characteristics of a distribution of differences in means. Instead, we can use what statisticians discovered long ago with the Central Limit Theorem: as long as the samples are independent, the sampling distribution for the *difference in means* can be described as follows.

1. The sampling distribution is centered around  $\mu_1 - \mu_2$ .
2. The standard deviation of our sampling distribution (i.e., standard error) is given as follows:

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

3. The shape of the sampling distribution is approximately normal if (1) both sample sizes are "sufficiently large" OR (2) if both original populations are normally distributed.

Given these characteristics, the test statistic we will use when testing for differences in two population means for INDEPENDENT samples is as follows:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

When the null hypothesis is true, this test statistic follows a t-distribution with the following degrees of freedom:

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

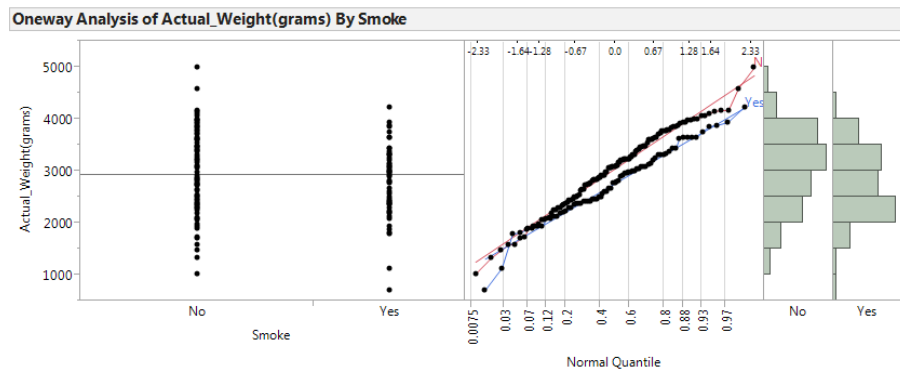
Therefore, as long as it is reasonable to assume the shape of the distribution of differences in means is approximately normal (see #3 above), we can use the t-distribution to find our p-value. This procedure is typically referred to as the **two-sample t-test**.

**Testing for a Difference in Two Population Means (Assuming Independent Samples)**

**Back to Example 6.5:** Carry out the hypothesis test associated with the following research question: Is the average birth weight lower when mothers are classified as smokers?

**Step 0:** Check the assumptions behind the test to be sure that the t-test is valid.

1. Are the two groups independent? Explain.
  
2. Are both sample sizes sufficiently large? If not, is it reasonable to assume that both populations are normally distributed?



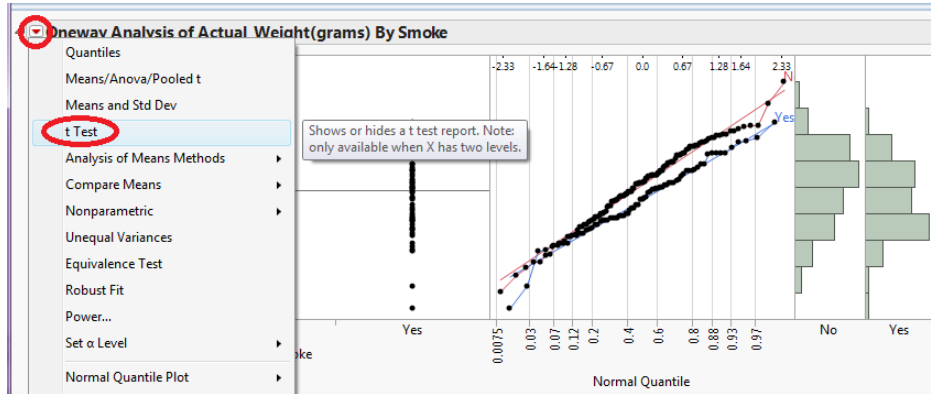
**Step 1:** Convert the research question into  $H_0$  and  $H_a$ .

$H_0$ :

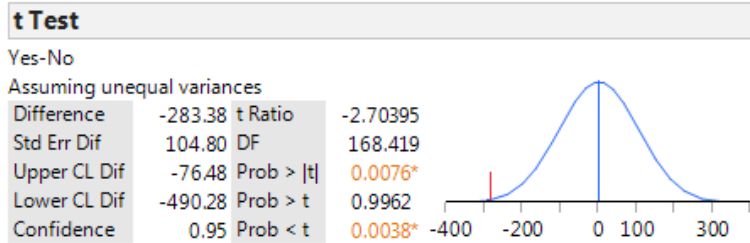
$H_a$ :

**Step 2:** Calculate a test statistic from your data.

To do this in JMP, click on the red drop-down arrow next to “Oneway Analysis...” and select **t-test**. You can also request that JMP display means and standard deviations for both groups.



JMP returns the following:



**Means and Std Deviations**

Level	Number	Mean	Std Dev	Std Err		
				Mean	Lower 95%	Upper 95%
No	113	3045.68	755.802	71.100	2904.8	3186.6
Yes	73	2762.30	657.851	76.996	2608.8	2915.8

First, identify the observed difference in the sample means:

$$\bar{x}_1 - \bar{x}_2 = \bar{x}_{\text{yes}} - \bar{x}_{\text{no}} =$$

Next, identify the standard error of the difference in means:

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} =$$

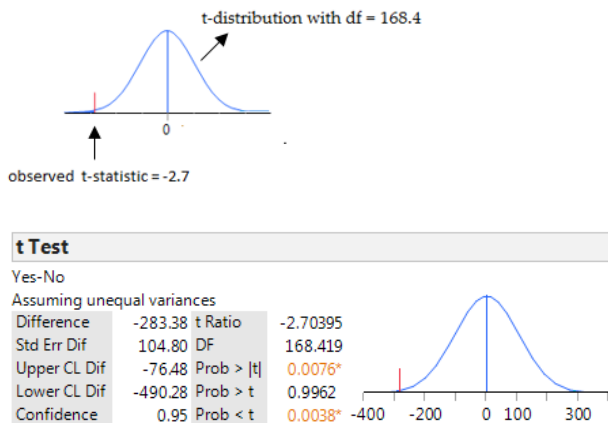
The test statistic will help us measure how far this observed difference in means is from what was hypothesized under the null hypothesis. JMP has already computed this test statistic, but we can verify the calculations as follows.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} =$$

Recall that this test statistic comes from a t-distribution with

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right)} = 168.419$$

**Step 3:** Determine the p-value.



p-value:

**Step 4:** Write a conclusion in terms of the original research question.

“We have evidence that the average birth weight is lower when mothers are classified as smokers (p-value = 0.0038).”

**Constructing a Confidence Interval for the Difference in Means**

Recall that the general form of a confidence interval is as follows:

General form of a confidence interval
sample statistic of interest $\pm$ margin of error

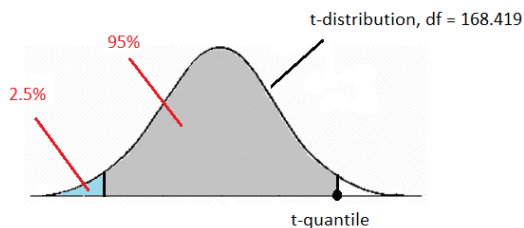
When comparing the means between two groups, the sample statistic of interest is:

$$\bar{x}_1 - \bar{x}_2 = 2762.30 - 3045.68 = -283.38$$

Like always, the margin of error is found by taking “about two” standard errors. Specifically, the margin of error is given by:

$$t\text{-quantile} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = t\text{-quantile} \sqrt{\frac{755.802^2}{113} + \frac{657.851^2}{73}} = t\text{-quantile} \times 104.80$$

The appropriate t-quantile can be found using JMP. Recall that for these data, we are using a t-distribution with  $df = 168.419$ . For a 95% confidence interval, the t-quantile is the t-value that separates the lower 97.5% of the t-distribution from the rest.



Open a new JMP data table and create a new column. Choose **Probability > t Quantile** from the Formula editor and enter the following:

```
t Quantile (0.975, 168.419, .)
```

For this example, JMP returns the following t-quantile:

```
t-quantile
1.9741495602
```



The 95% confidence interval can then be computed as follows:

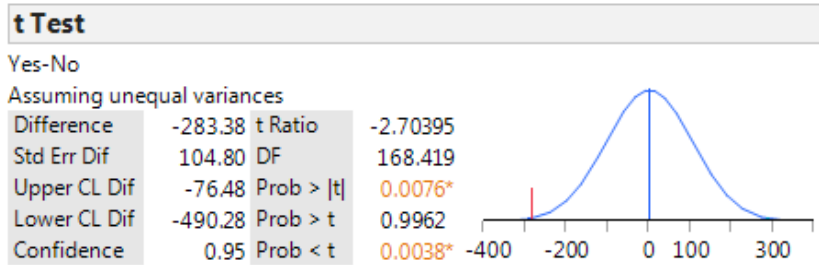
$$(\bar{x}_1 - \bar{x}_2) \pm t\text{-quantile} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \text{ which is}$$

$$-283.38 \pm 1.97415 (104.80), \text{ or}$$

$$-283.38 \pm 206.89$$

This 95% confidence interval for the true difference in population means can be written as  $-490.3 \leq \mu_1 - \mu_2 \leq -76.5$ , or  $-490.3 \leq \mu_{\text{Yes}} - \mu_{\text{No}} \leq -76.5$ .

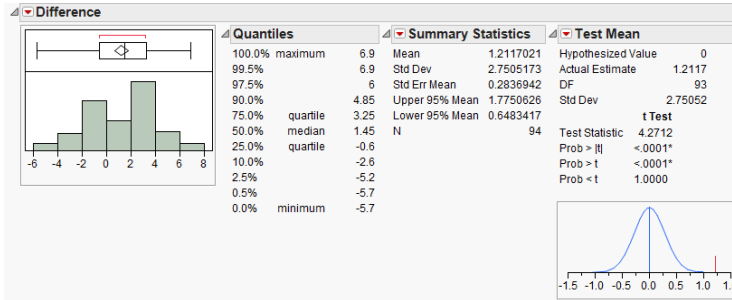
Note that these endpoints are also given in the JMP output for the two-sample t-test.



Interpret the meaning of this 95% confidence interval:

**Example 6.6: The Importance of Using a Paired t-test When Samples are Dependent**

Recall the Thanksgiving break weight gain data from Example 6.1. Recall that these data were paired since each subject was measured twice, so we used a paired t-test for dependent samples.



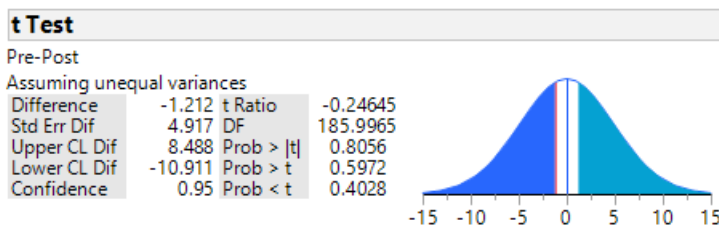
What if the researchers had mistakenly treated the observations as independent and used a two-sample t-test?

The Data:

	Time	Weight
1	Pre	149.7
2	Pre	119.5
3	Pre	166.5
4	Pre	198.3
5	Pre	195.9
...	...	...
185	Post	174.2
186	Post	176.4
187	Post	166.3
188	Post	129.6

Analyze > Fit Y by X:

Results:



Note that the results of the two-sample t-test does not detect a significant difference between the mean pre-weight and the mean post-weight; the paired t-test, however, did reveal a significant weight gain. Why? Notice that there is a lot of variation in weights from one subject to the next. If the difference between time periods (i.e., pre vs. post) is small relative to the differences between subjects, then it would take a very large sample size to get a significant two-sample t-test result. A paired t-test, however, just looks at the differences between time periods and is more powerful than the two-sample t-test.

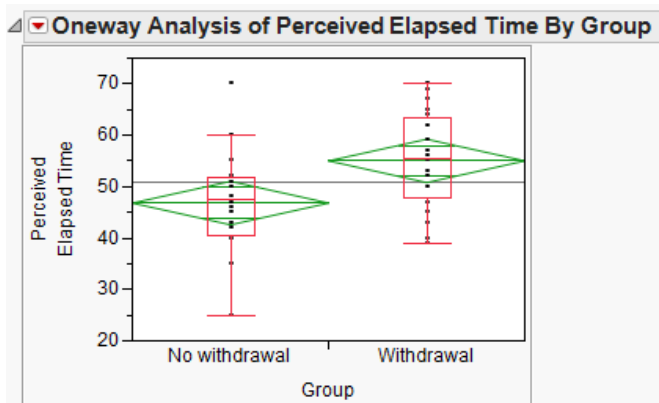
**Example 6.7:** Recall the study conducted by researchers at Pennsylvania State University which investigated whether time perception, a simple indication of a person's ability to concentrate, is impaired during nicotine withdrawal. This study was discussed in Chapter 5. Twenty smokers were randomly assigned to a 24-hour smoking abstinence and were asked to estimate how much time had passed during a 45-second period. Another 20 smokers were randomly assigned to a group which was *not* forced to abstain from nicotine; they were also asked to estimate how much time had passed during a 45-second period. Suppose the resulting data on perceived elapsed time (in seconds) were analyzed as shown below (these results are artificial but are similar to the actual findings). The data can be found in the file **Nicotine\_all.JMP**.

Research Question: Do those smokers suffering from nicotine withdrawal tend to believe that more time has elapsed than do those not suffering from nicotine withdrawal?

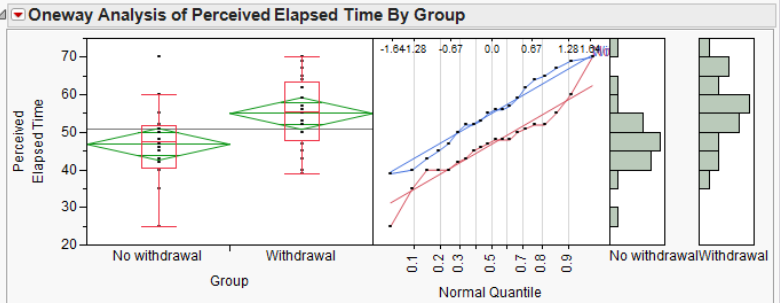
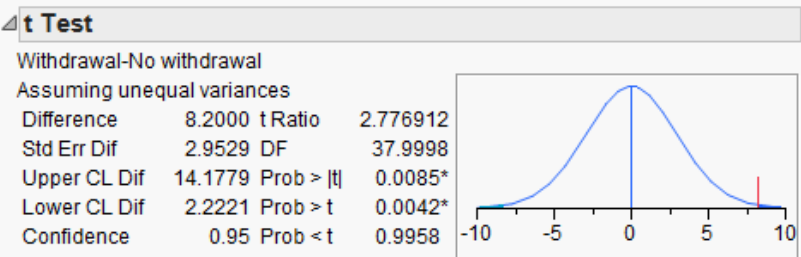
Questions:

1. What is the response variable in this study?
2. What is the predictor variable in this study?

The data can be summarized using JMP by selecting **Analyze > Fit Y by X** and placing the predictor (i.e., explanatory) variable in the **X, Factor** box and the response in the **Y, Response** box. Then, use the **Display Options** menu from the red drop-down arrow to add the boxplots and mean diamonds.



It is clear that the SAMPLE of smokers suffering from nicotine withdrawal has a larger mean perceived elapsed time than those not suffering from withdrawal. However, to determine whether there is a difference in the POPULATION means, we will carry out a hypothesis test.

<p><b>Step 0:</b></p>	<p>Check the assumptions behind the test to be sure that the test is valid.</p> <ol style="list-style-type: none"> <li>1. Are the two groups independent?</li> <li>2. Are both sample sizes sufficiently large? If not, is it reasonable to assume that both populations are normally distributed?</li> </ol> 																																
<p><b>Step 1:</b></p>	<p>Convert the research question into <math>H_0</math> and <math>H_a</math>.</p> <p><math>H_0</math>:</p> <p><math>H_a</math>:</p>																																
<p><b>Step 2:</b></p>	<p>Calculate a <u>test statistic</u> from your data.</p>  <table border="1"> <thead> <tr> <th colspan="4">t Test</th> </tr> </thead> <tbody> <tr> <td colspan="4">Withdrawal-No withdrawal</td> </tr> <tr> <td colspan="4">Assuming unequal variances</td> </tr> <tr> <td>Difference</td> <td>8.2000</td> <td>t Ratio</td> <td>2.776912</td> </tr> <tr> <td>Std Err Dif</td> <td>2.9529</td> <td>DF</td> <td>37.9998</td> </tr> <tr> <td>Upper CL Dif</td> <td>14.1779</td> <td>Prob &gt;  t </td> <td>0.0085*</td> </tr> <tr> <td>Lower CL Dif</td> <td>2.2221</td> <td>Prob &gt; t</td> <td>0.0042*</td> </tr> <tr> <td>Confidence</td> <td>0.95</td> <td>Prob &lt; t</td> <td>0.9958</td> </tr> </tbody> </table>	t Test				Withdrawal-No withdrawal				Assuming unequal variances				Difference	8.2000	t Ratio	2.776912	Std Err Dif	2.9529	DF	37.9998	Upper CL Dif	14.1779	Prob >  t	0.0085*	Lower CL Dif	2.2221	Prob > t	0.0042*	Confidence	0.95	Prob < t	0.9958
t Test																																	
Withdrawal-No withdrawal																																	
Assuming unequal variances																																	
Difference	8.2000	t Ratio	2.776912																														
Std Err Dif	2.9529	DF	37.9998																														
Upper CL Dif	14.1779	Prob >  t	0.0085*																														
Lower CL Dif	2.2221	Prob > t	0.0042*																														
Confidence	0.95	Prob < t	0.9958																														

Verify the value of the test statistic from JMP:

Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err		
				Mean	Lower 95%	Upper 95%
No withdrawal	20	46.8500	9.34866	2.0904	42.475	51.225
Withdrawal	20	55.0500	9.32724	2.0856	50.685	59.415

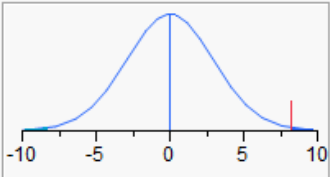
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} =$$

Recall that this test statistic comes from a t-distribution with

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right)} = 37.9998$$

**Step 3:** Determine the p-value.

t Test			
Withdrawal-No withdrawal			
Assuming unequal variances			
Difference	8.2000	t Ratio	2.776912
Std Err Dif	2.9529	DF	37.9998
Upper CL Dif	14.1779	Prob >  t	0.0085*
Lower CL Dif	2.2221	Prob > t	0.0042*
Confidence	0.95	Prob < t	0.9958

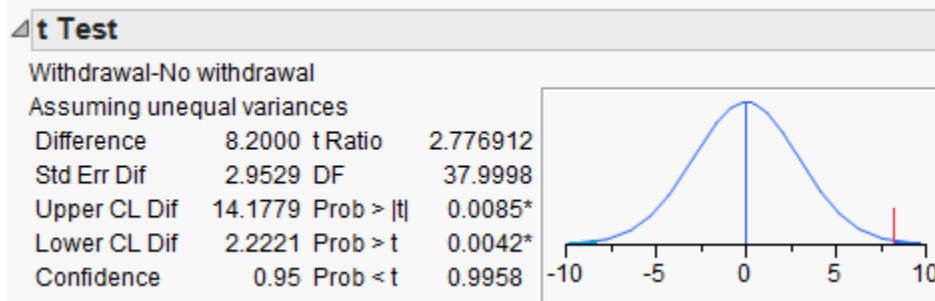


p-value:

**Step 5:** Write a conclusion in terms of the original research question.

“The study provides evidence that the average perceived elapsed time is higher for those smokers suffering from nicotine withdrawal than for those not suffering from nicotine withdrawal (p-value = .0042).”

Finally, construct a 95% confidence interval for the difference in means:



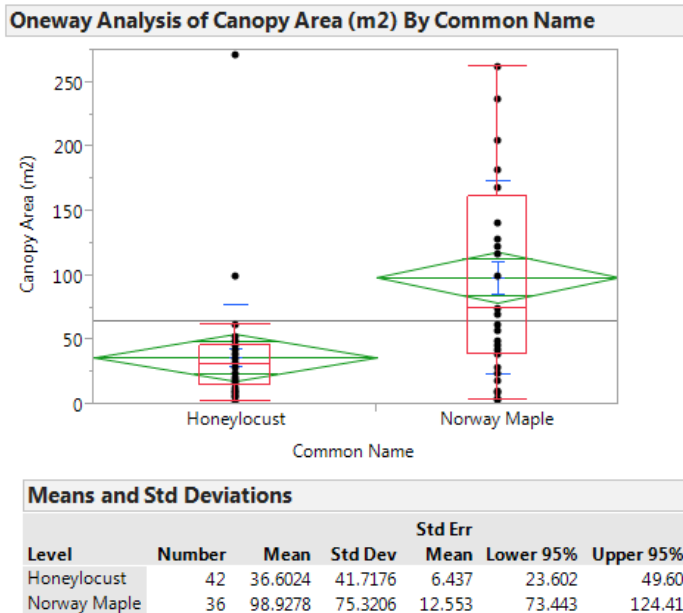
Lower Endpoint =

Upper Endpoint =

Again, we often write this confidence interval as:

Interpret the meaning of this confidence interval:

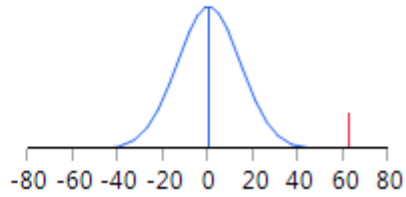
**Example 6.8:** Consider the data found in the file **NYC\_Trees.JMP**. Suppose you want to compare the canopy area of the two types of trees. Is the average canopy area of a Honey Locust different from the average canopy area of a Norway Maple? The data are summarized as follows:



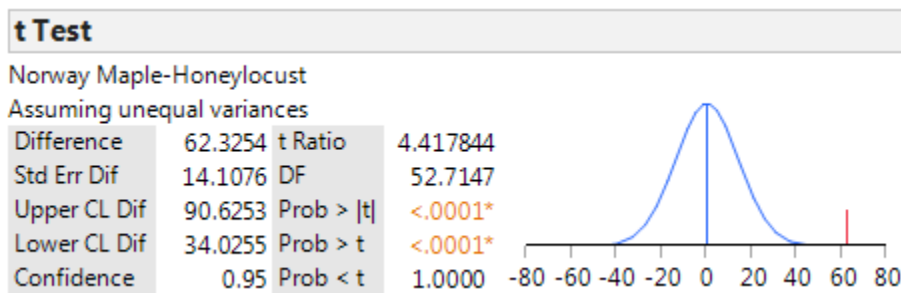
It is clear that the SAMPLE of Norway Maple trees has a higher mean canopy area than the SAMPLE of Honey Locust trees. However, to determine whether the POPULATION mean canopy areas differ between the two groups, we will carry out a hypothesis test.

**Step 0:** Check the assumptions behind the test to be sure that the test is valid.

1. Are the two groups independent?
2. Are both sample sizes sufficiently large? If not, is it reasonable to assume that both populations are normally distributed?

<p><b>Step 1:</b></p>	<p>Convert the research question into <math>H_0</math> and <math>H_a</math>.</p> <p><math>H_0</math>:</p> <p><math>H_a</math>:</p>																				
<p><b>Step 2:</b></p>	<p>Calculate a <u>test statistic</u> from your data.</p> <div data-bbox="386 615 1279 898" style="border: 1px solid gray; padding: 5px;"> <p><b>t Test</b></p> <p>Norway Maple-Honeylocust</p> <p>Assuming unequal variances</p> <table border="1"> <tr> <td>Difference</td> <td>62.3254</td> <td>t Ratio</td> <td>4.417844</td> </tr> <tr> <td>Std Err Dif</td> <td>14.1076</td> <td>DF</td> <td>52.7147</td> </tr> <tr> <td>Upper CL Dif</td> <td>90.6253</td> <td>Prob &gt;  t </td> <td>&lt;.0001*</td> </tr> <tr> <td>Lower CL Dif</td> <td>34.0255</td> <td>Prob &gt; t</td> <td>&lt;.0001*</td> </tr> <tr> <td>Confidence</td> <td>0.95</td> <td>Prob &lt; t</td> <td>1.0000</td> </tr> </table>  </div>	Difference	62.3254	t Ratio	4.417844	Std Err Dif	14.1076	DF	52.7147	Upper CL Dif	90.6253	Prob >  t	<.0001*	Lower CL Dif	34.0255	Prob > t	<.0001*	Confidence	0.95	Prob < t	1.0000
Difference	62.3254	t Ratio	4.417844																		
Std Err Dif	14.1076	DF	52.7147																		
Upper CL Dif	90.6253	Prob >  t	<.0001*																		
Lower CL Dif	34.0255	Prob > t	<.0001*																		
Confidence	0.95	Prob < t	1.0000																		
<p><b>Step 3:</b></p>	<p>Determine the <u>p-value</u>.</p> <p><u>p-value</u>:</p>																				
<p><b>Step 4:</b></p>	<p>Write a conclusion in terms of the original research question.</p>																				

Finally, we can construct a confidence interval for the difference in means:





Lower Endpoint =

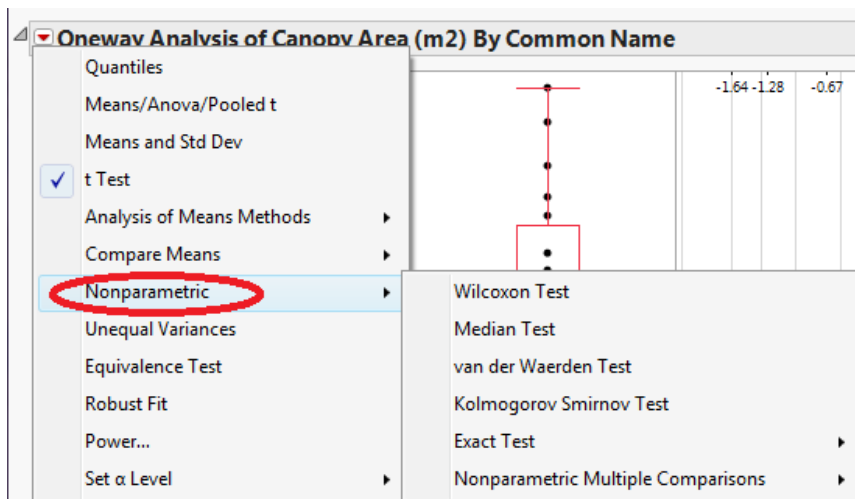
Upper Endpoint =

This confidence interval can also be written as:

Interpret the meaning of this 95% confidence interval:

### Nonparametric Tests

These tests are an alternative to the two-sample t-test for comparing the “average” value of two populations where the samples from each population are taken independently. This test is appropriate when the assumption of normality is violated (as it may have been in the previous example). To obtain the output for these tests, simply select **Nonparametric** > from the drop-down menu in the upper left-hand corner of the boxplots.



JMP returns the following output when the Wilcoxon Test is selected:

<b>Wilcoxon / Kruskal-Wallis Tests (Rank Sums)</b>					
Level	Count	Score Sum	Expected		
			Score	Score Mean	(Mean-Mean0)/Std0
Honeylocust	42	1252.00	1659.00	29.8095	-4.080
Norway Maple	36	1829.00	1422.00	50.8056	4.080

<b>2-Sample Test, Normal Approximation</b>		
S	Z	Prob> Z
1829	4.08024	<.0001*

<b>1-way Test, ChiSquare Approximation</b>		
ChiSquare	DF	Prob>ChiSq
16.6894	1	<.0001*

The decision based on Wilcoxon’s test is the same decision that we arrived at using the standard two-sample t-test above. Often, these procedures will agree each other because the above t-test is robust to (i.e. not greatly affected by) departures from normality.

### The t-Test for Independent Samples Assuming Equal Variances

The two-sample t-test that we discussed in this chapter uses the following test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Many statisticians use this version of the t-test only when they feel that the variances of the two populations of interest are significantly different from one another (hence it is often called the two-sample t-test assuming unequal variances).

If, however, it is reasonable to assume that the amount of variability in the two populations is about the same, then statisticians often use a slightly different version of the t-test. Assuming the two variances are equal, we can obtain a common estimate of this equal variance (this is called a *pooled* estimate).

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

Then, the test statistic for the t-test is calculated as follows:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_{\text{pooled}}^2}{n_1} + \frac{s_{\text{pooled}}^2}{n_2}}}$$

Assuming the two variances are equal, this test statistic comes from a t-distribution with  $df = (n_1 - 1) + (n_2 - 1)$ .

In practice, the two-sample t-test that was introduced in this chapter works well, even if the two distributions being compared have the same amount of variability. Therefore, it's my opinion that instead of worrying about which one to use, just go with the test assuming unequal variances.

Note, however, that if you continue studying statistics, you might hear a reference to a two-sample t-test assuming equal variances.