In this chapter, we will consider methods that allow us to make comparisons on numerical variables between two different groups. In general, these methods should be used to address research questions involving a categorical predictor variable (with two categories) and a numerical response variable. Two different cases will be considered: (1) When the samples are *dependent* (also called *matched* or *paired*), and (2) when the samples are independent.

COMPARING TWO POPULATION MEANS: DEPENDENT SAMPLES

The hypothesis testing procedures presented in this section should be used when the observations from the two groups being compared are <u>dependent</u>. Whether the observations are dependent is determined by how the data are collected – if each observational unit is measured twice, then the observations are dependent (i.e., matched or paired). To see this, consider the following example.

Example 6.1: Thanksgiving Holiday Weight Gain

In a 2006 study published in *Nutrition Journal*, researchers were interested in assessing potential changes that occur in body weight during the Thanksgiving holiday break in college students. This paper's abstract indicates that "a total of 94 college students reported to the human body composition laboratory at the University of Oklahoma following a 6-hour fast with testing occurring prior to, and immediately following the Thanksgiving holiday break. Body weight was assessed using a balance beam scale while participants were dressed in minimal clothing."
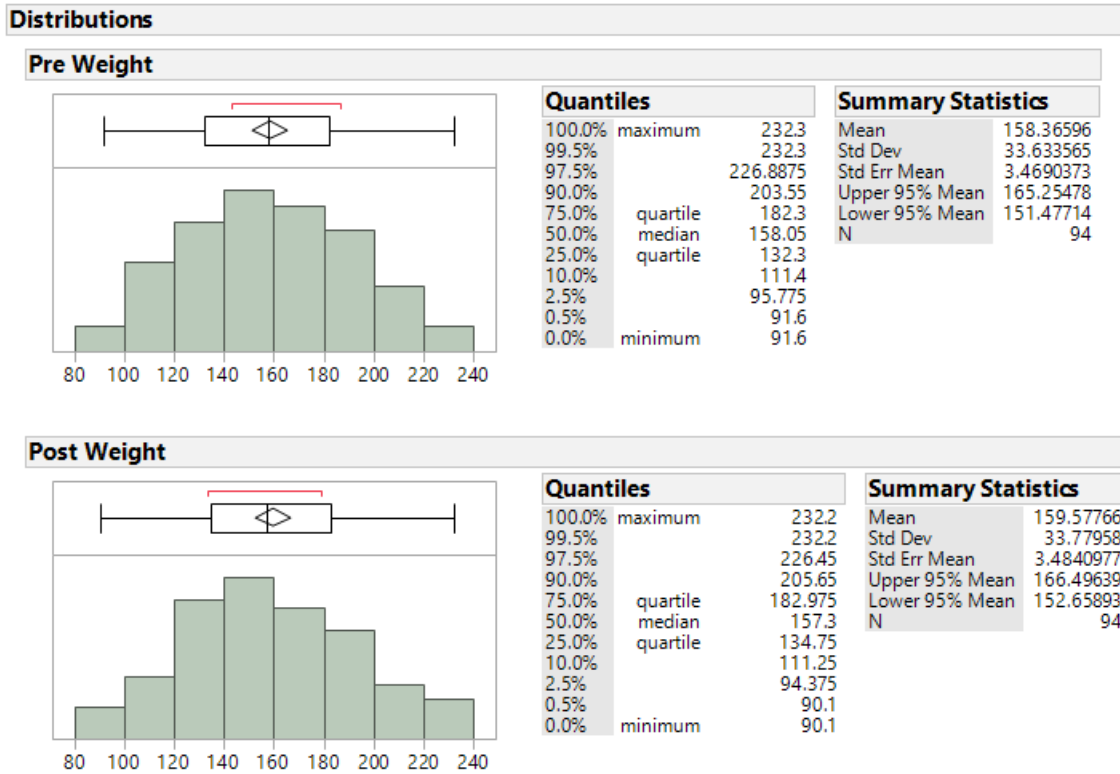
*Source: Hull et al. The Effects of the Thanksgiving Holiday on Weight Gain. Nutrition Journal 2006, **5**:29; doi:10.1186/1475-2891-5-29.*

<u>Research Question</u>: On average, does the body weight of college students increase over the Thanksgiving holiday break?

The raw data from this study were not available; however, the data in the file **ThanksgivingWeightData.jmp** are similar to the actual results obtained by the researchers. A portion of these data is shown below:

	Pre Weight	Post Weight
1	149.7	153.6
2	119.5	122.1
3	166.5	166.2
4	198.3	198.1
5	195.9	196.5
6	212.4	210.5
7	91.6	90.1
.	.	.
.	.	.
.	.	.
92	178	176.4
93	165.5	166.3
94	127.3	129.6

For illustration purposes only, let's analyze the distribution of the pre- and post-weights separately.

Distributions

Pre Weight

Quantiles

100.0%	maximum	232.3
99.5%		232.3
97.5%		226.8875
90.0%		203.55
75.0%	quartile	182.3
50.0%	median	158.05
25.0%	quartile	132.3
10.0%		111.4
2.5%		95.775
0.5%		91.6
0.0%	minimum	91.6

Summary Statistics

Mean	158.36596
Std Dev	33.633565
Std Err Mean	3.4690373
Upper 95% Mean	165.25478
Lower 95% Mean	151.47714
N	94

Post Weight

Quantiles

100.0%	maximum	232.2
99.5%		232.2
97.5%		226.45
90.0%		205.65
75.0%	quartile	182.975
50.0%	median	157.3
25.0%	quartile	134.75
10.0%		111.25
2.5%		94.375
0.5%		90.1
0.0%	minimum	90.1

Summary Statistics

Mean	159.57766
Std Dev	33.77958
Std Err Mean	3.4840977
Upper 95% Mean	166.49639
Lower 95% Mean	152.65893
N	94

Questions:

1. How do the mean weights compare from before to after Thanksgiving break?

2. Consider the confidence intervals for the mean weight at each time period. What conclusion might you draw from these confidence intervals?

Comment: One of the problems with the above analysis approach (i.e., examining the two confidence intervals) is that it is treating the observations from the two time periods as if they are independent, or not related. For these data, however, the first pre-weight is clearly related to the first post-weight (the two measurements were made on the same person). Thus, these two samples are **dependent**.

In other words, some of the variability in the weight measurements might be due to differences between time periods (before vs. after Thanksgiving), but much of the variability in the observations is likely due to differences between people!

So, to control for the variability in weights from person to person (which will help us isolate the true differences between pre- and post-Thanksgiving break), we will work with the DIFFERENCES on each subject, instead. This will remove the structure of dependence between the pre- and post-Thanksgiving groups and will control for the fact that some people, in general, tend to weigh more (or less) than others. In the end, this helps us to isolate the differences in weight measurements that can be attributed to the effect of before vs. after break.

Calculating the Differences in JMP

First, open the file called **ThanksgivingWeightData.** To calculate the differences, create an additional column (double click on empty column next to "'Post Weight") and title it "Difference." Right-click on the new column header and select **Formula**.
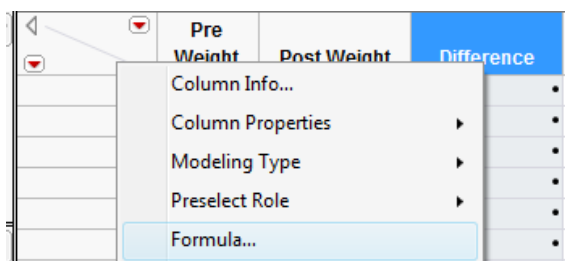
In the edit window, tell JMP to calculate the difference as follows:
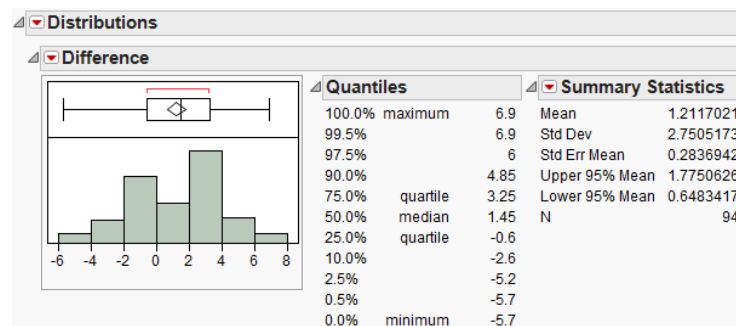

Click **Apply** and then **OK**, and JMP returns the following (only the results for the first 10 subjects are shown below):

	Pre Weight	Post Weight	Difference
1	149.7	153.6	3.9
2	119.5	122.1	2.6
3	166.5	166.2	-0.3
4	198.3	198.1	-0.2
5	195.9	196.5	0.6
6	212.4	210.5	-1.9
7	91.6	90.1	-1.5
8	151.7	154.9	3.2
9	192.7	194.1	1.4
10	125.4	128.8	3.4

Questions:

1. What does a positive difference of 3.9 indicate?

2. What does a difference of -1.9 indicate?

To determine what the average difference is, we can analyze the distribution of the differences using the descriptive methods discussed in Chapter 4:

Distributions

Difference

Quantiles				Summary Statistics	
100.0%	maximum	6.9		Mean	1.2117021
99.5%		6.9		Std Dev	2.7505173
97.5%		6		Std Err Mean	0.2836942
90.0%		4.85		Upper 95% Mean	1.7750626
75.0%	quartile	3.25		Lower 95% Mean	0.6483417
50.0%	median	1.45		N	94
25.0%	quartile	-0.6			
10.0%		-2.6			
2.5%		-5.2			
0.5%		-5.7			
0.0%	minimum	-5.7			

Questions:

3. What is the average difference of the 94 subjects? Interpret this value.

4. If Thanksgiving break had no effect on body weight, what would you expect these differences to be, on average?

5. If the research hypothesis is correct (body weights of college students increase over Thanksgiving break), what would you expect these differences to be, on average?

Comment: Note that these differences are represented by a single column of data. *So, you could view this as a problem involving a single numerical variable – the differences!* Therefore, the hypothesis testing procedure is *exactly the same* as the procedure for testing claims about a single population mean we discussed in Chapter 5.

That is, the parameter of interest is the true population average of the differences which we will represent by $\mu_{difference}$.

- Our best estimate for this parameter is the <u>sample mean</u> of the observed differences. We'll call this quantity $\overline{x}_{difference}$.

- The sample standard deviation of the differences will be denoted by $s_{difference}$.

We can carry out the hypothesis test as follows to determine whether body weights of college students tend to increase over Thanksgiving break.

<u>Checking the Normality Assumption:</u>

Recall that for the t-test to be valid, at least one of the following conditions must be met:

- Either the sample size (i.e., number of pairs) is sufficiently large (greater than about 30), or
- The distribution of the observed differences is approximately normal (which would indicate that the original population is normally distributed so that we can be certain the distribution of sample means is normally distributed even with a small sample size)

Question: Does the t-test appear to be a valid approach for testing this research question? Justify your reasoning.

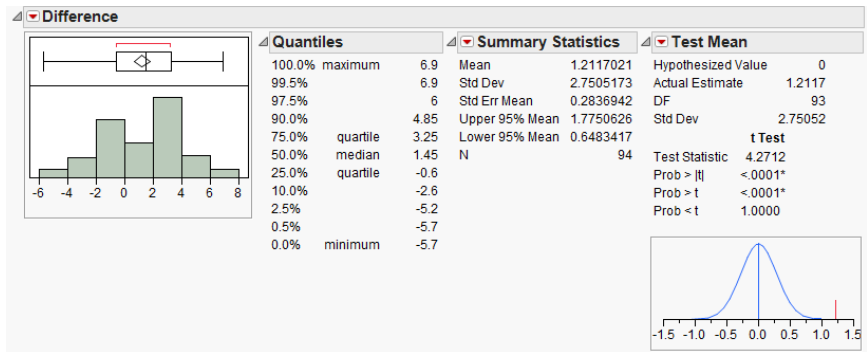**<u>Step 1: Set up the null and alternative hypotheses</u>**

H$_o$:

H$_a$:

Step 2: Find the t-statistic and the p-value

$$t = \frac{\overline{x}_{difference} - \mu_{difference}}{s_{difference}/\sqrt{n}} =$$

To calculate this test statistic and its associated p-value in JMP, use the methods discussed in Chapter 5:

Difference

Quantiles			Summary Statistics		Test Mean	
100.0%	maximum	6.9	Mean	1.2117021	Hypothesized Value	0
99.5%		6.9	Std Dev	2.7505173	Actual Estimate	1.2117
97.5%		6	Std Err Mean	0.2836942	DF	93
90.0%		4.85	Upper 95% Mean	1.7750626	Std Dev	2.75052
75.0%	quartile	3.25	Lower 95% Mean	0.6483417		t Test
50.0%	median	1.45	N	94		
25.0%	quartile	-0.6			Test Statistic	4.2712
10.0%		-2.6			Prob > |t|	<.0001*
2.5%		-5.2			Prob > t	<.0001*
0.5%		-5.7			Prob < t	1.0000
0.0%	minimum	-5.7				

p-value =

Step 3: Write a conclusion in the context of the problem

Interpreting the 95% confidence interval for the mean difference

Comment: Suppose that when calculating the difference, you switched the order of the categories:
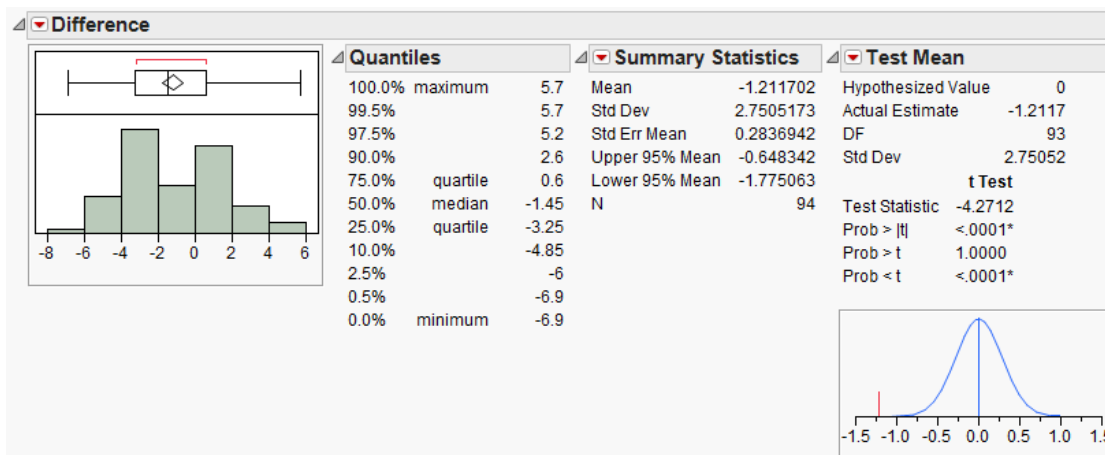
Pre Weight - Post Weight

What effect would this have had on the analysis?

Ho:

Ha:

$$t = \dfrac{\overline{X}_{difference} - \mu_{difference}}{S_{difference}\Big/\sqrt{n}} =$$

Difference

Quantiles		
100.0%	maximum	5.7
99.5%		5.7
97.5%		5.2
90.0%		2.6
75.0%	quartile	0.6
50.0%	median	-1.45
25.0%	quartile	-3.25
10.0%		-4.85
2.5%		-6
0.5%		-6.9
0.0%	minimum	-6.9

Summary Statistics	
Mean	-1.211702
Std Dev	2.7505173
Std Err Mean	0.2836942
Upper 95% Mean	-0.648342
Lower 95% Mean	-1.775063
N	94

Test Mean	
Hypothesized Value	0
Actual Estimate	-1.2117
DF	93
Std Dev	2.75052

t Test	
Test Statistic	-4.2712
Prob > |t|	<.0001*
Prob > t	1.0000
Prob < t	<.0001*

p-value =

Conclusion:

95% confidence interval for the mean difference:

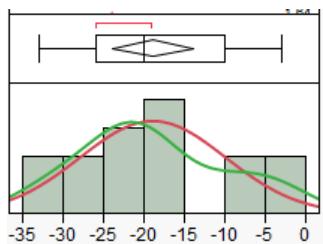**Example 6.2: Systolic Blood Pressure and Captopril**

The data in the file **Captopril.jmp** give the systolic and diastolic blood pressures for 15 patients with moderate essential hypertension, immediately before and two hours after taking a drug, captopril. Our interest is in investigating the response to the drug treatment.

Research Question: Is there a change in systolic blood pressure after taking captopril?

Question: Are these samples dependent or independent? Explain.

Check the assumptions behind the test

- Is the number of pairs sufficiently large?
- If not, is it reasonable to assume the differences in systolic and diastolic blood pressures are normally distributed?
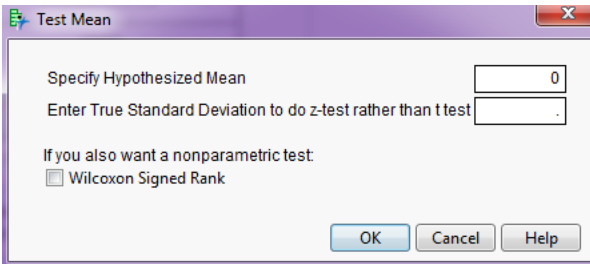
Step 1: Set up the null and alternative hypotheses

H$_0$:

H$_a$:

Step 2: Find the test statistic and the p-value

Summary Statistics	
Mean	-18.93333
Std Dev	9.0274713
Std Err Mean	2.3308831
Upper 95% Mean	-13.93409
Lower 95% Mean	-23.93258
N	15

$$t = \frac{\overline{x}_{difference} - \mu_{difference}}{s_{difference} / \sqrt{n}} =$$

Using JMP:

Test Mean		
Hypothesized Value		0
Actual Estimate		-18.933
DF		14
Std Dev		9.02747
t Test		
Test Statistic	-8.1228	
Prob > |t|	<.0001*	
Prob > t	1.0000	
Prob < t	<.0001*	
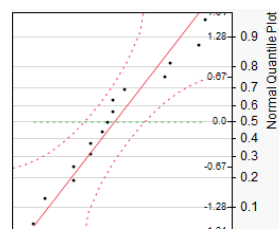
p-value:

Step 3: Write a conclusion in the context of the problem

Finally, construct a 95% confidence interval for the average difference in systolic blood pressure means.

Summary Statistics	
Mean	-18.93333
Std Dev	9.0274713
Std Err Mean	2.3308831
Upper 95% Mean	-13.93409
Lower 95% Mean	-23.93258
N	15

Questions:

1. Interpret this confidence interval.

2. Does this interval agree with the results of the hypothesis test? Explain.
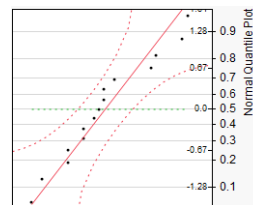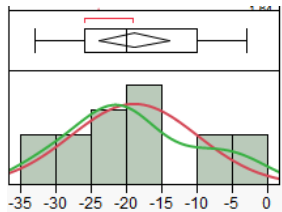
Example 6.3: Systolic Blood Pressure and Captopril, Revisited

As we saw in Example 6.2, patients with moderate essential hypertension saw a decrease in systolic blood pressure. Note that we could also investigate the following question.

Research Question: Does the systolic blood pressure decrease by more than 10 mmHg?

Check the assumptions behind the test

- Is the number of pairs sufficiently large?
- If not, is it reasonable to assume the differences in systolic and diastolic blood pressures are normally distributed?
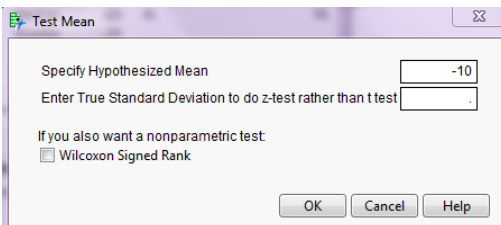
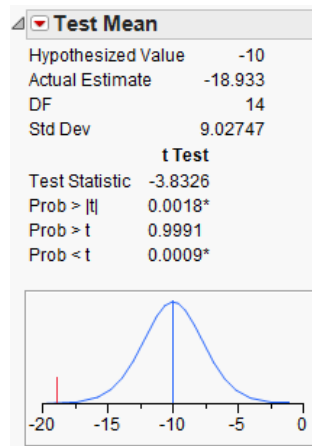**Step 1: Set up the null and alternative hypotheses**

H₀:

Hₐ:

Step 2: Find the test statistic and the p-value

Summary Statistics	
Mean	-18.93333
Std Dev	9.0274713
Std Err Mean	2.3308831
Upper 95% Mean	-13.93409
Lower 95% Mean	-23.93258
N	15

$$t = \frac{\overline{X}_{difference} - \mu_{difference}}{S_{difference}\big/\sqrt{n}} =$$

Using JMP:

Test Mean

Specify Hypothesized Mean   -10

Enter True Standard Deviation to do z-test rather than t test

If you also want a nonparametric test:
☐ Wilcoxon Signed Rank

OK   Cancel   Help

Test Mean

Hypothesized Value	-10
Actual Estimate	-18.933
DF	14
Std Dev	9.02747

t Test

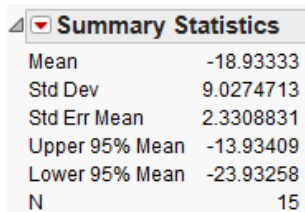|	
Test Statistic	-3.8326
Prob > |t|	0.0018*
Prob > t	0.9991
Prob < t	0.0009*

p-value:

Step 3: Write a conclusion in the context of the problem

Recall the 95% confidence interval for the average difference in systolic blood pressure means:

Summary Statistics

Mean	-18.93333
Std Dev	9.0274713
Std Err Mean	2.3308831
Upper 95% Mean	-13.93409
Lower 95% Mean	-23.93258
N	15

Does this interval agree with the results of the hypothesis test? Explain.

Example 6.4: Diastolic Blood Pressure and Captopril

Finally, we will use the Captopril data to investigate the following question.

Research Question: Does the diastolic blood pressure decrease by more than 5 mmHg?

Step 0: Check the assumptions behind the test

- Is the number of pairs sufficiently large?
- If not, is it reasonable to assume the differences in systolic and diastolic blood pressures
  are normally distributed?

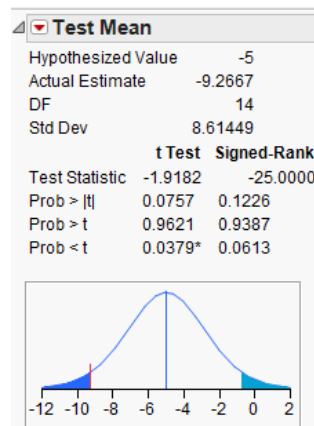### Step 1: Set up the null and alternative hypotheses

H₀:

Hₐ:
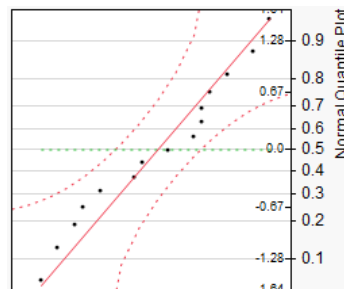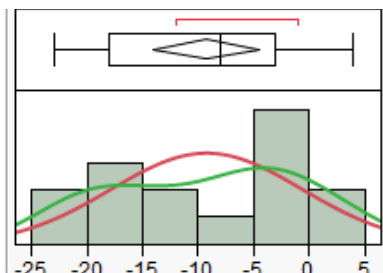
Step 2: Find the test statistic and the p-value

Summary Statistics

Mean	-9.266667
Std Dev	8.614495
Std Err Mean	2.224253
Upper 95% Mean	-4.496118
Lower 95% Mean	-14.03721
N	15

$$t = \frac{\overline{x}_{difference} - \mu_{difference}}{s_{difference}\big/\sqrt{n}} =$$

174

Using JMP:

Test Mean (dialog box)

Specify Hypothesized Mean   -5

Enter True Standard Deviation to do z-test rather than t test [.]

If you also want a nonparametric test:
☑ Wilcoxon Signed Rank

[OK] [Cancel] [Help]

Test Mean

Hypothesized Value	-5
Actual Estimate	-9.2667
DF	14
Std Dev	8.61449

	t Test	Signed-Rank
Test Statistic	-1.9182	-25.0000
Prob > |t|	0.0757	0.1226
Prob > t	0.9621	0.9387
Prob < t	0.0379*	0.0613

-12 -10 -8 -6 -4 -2 0 2
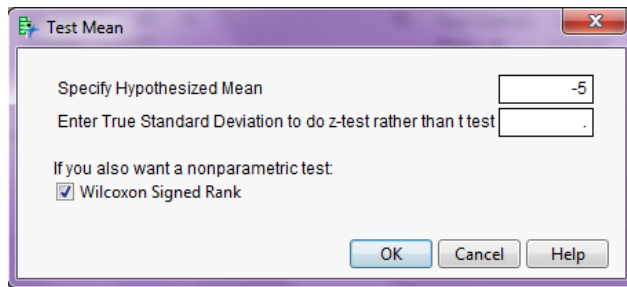
p-value:

Step 3: Write a conclusion in the context of the problem

Find the 95% confidence interval for the average difference in systolic blood pressure means:

Summary Statistics

Mean	-9.266667
Std Dev	8.614495
Std Err Mean	2.224253
Upper 95% Mean	-4.496118
Lower 95% Mean	-14.03721
N	15

Does this interval agree with the results of the hypothesis test? Explain.