Estimation and Hypothesis Testing: Two Populations

Chapter 10

Inferences Concerning Independent Samples: Known Variances, and Equal or Unequal Unknown Variances

It is important to note that Minitab does not offer a 2-sample *Z*-test. Instead, the 2-sample *t*-test is used to construct any inferences for $\mu_1 - \mu_2$.

Constructing a Confidence Interval for Independent Samples

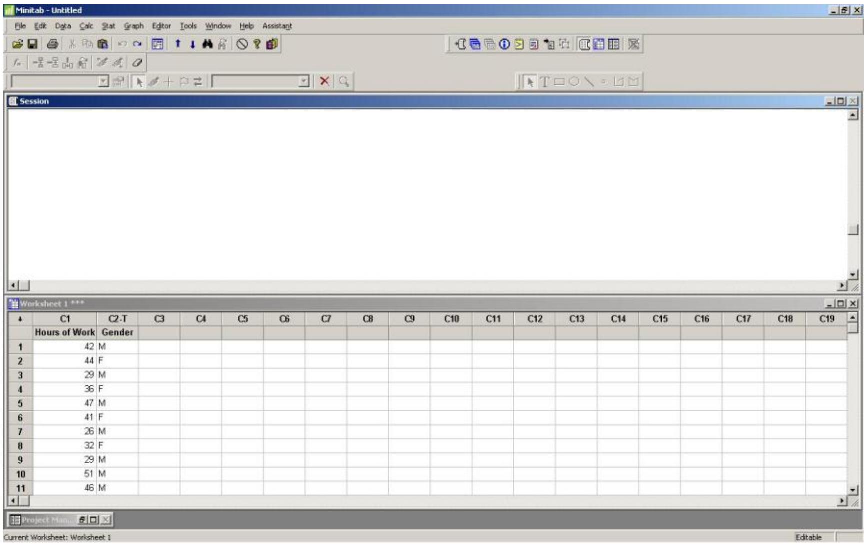
In a random sample of 25 men at a large corporation, the number of hours spent at work in the last week was collected.

42	51	42	36	49
29	46	37	33	41
47	41	28	39	48
26	35	37	48	39
29	31	44	38	46

In a random sample of 20 women at the same corporation, the number of hours spent at work in the last week was collected.

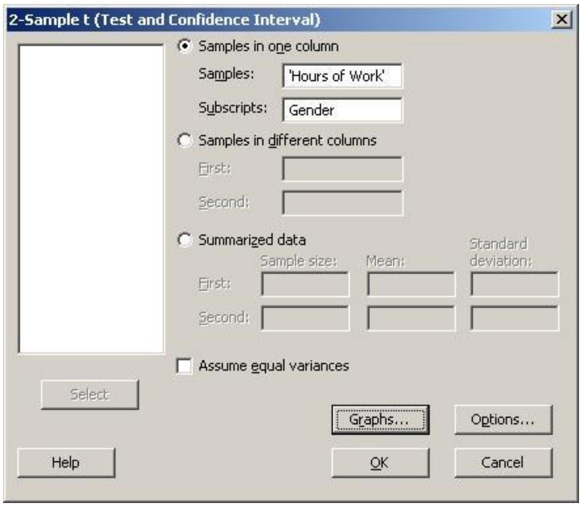
44	50	48	37	42
36	37	39	41	43
41	29	32	39	48
32	40	37	46	38

There are 3 ways to enter this data into Minitab to be analyzed. The first way is to enter the data into 1 column with an indicator in the second column as to the group each data point belongs. Assume that an indicator of "M" is for a data value of a Male, and "F" is an indicator for a female. Enter the data values in column C1 and label the column "Hours of Work." Enter the indicators in column C2 and label the column "Gender."

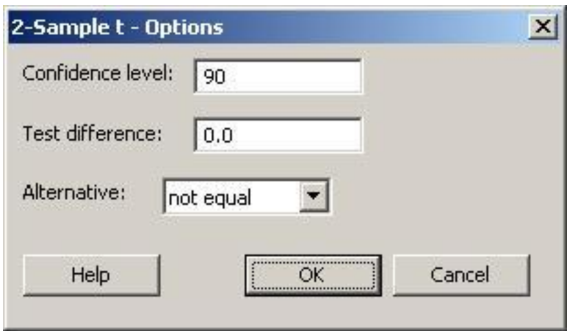


Management is interested in creating a confidence interval for the difference between the mean hours of work of men and women at their corporation.

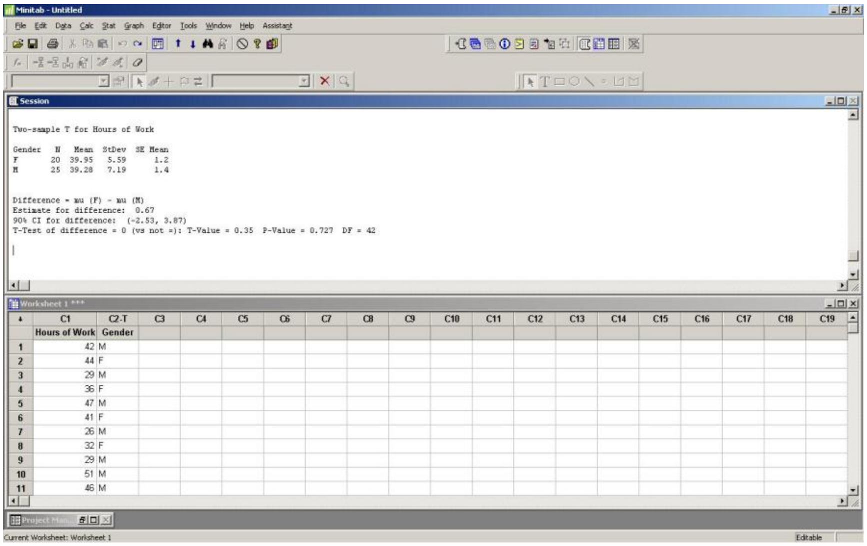
To construct a 90% confidence interval for the mean, click on **Stat → Basic Statistics → 2-Sample t.** Select the circle next to **Samples in one column.** Select the data in column C1 for the **Samples** field, and the data in column C2 for the **Subscripts** field. If you assume that the two populations have equal variances, select the **Assume equal variances** checkbox. If you do not have this assumption, do not select the checkbox. Select the **Options** button.



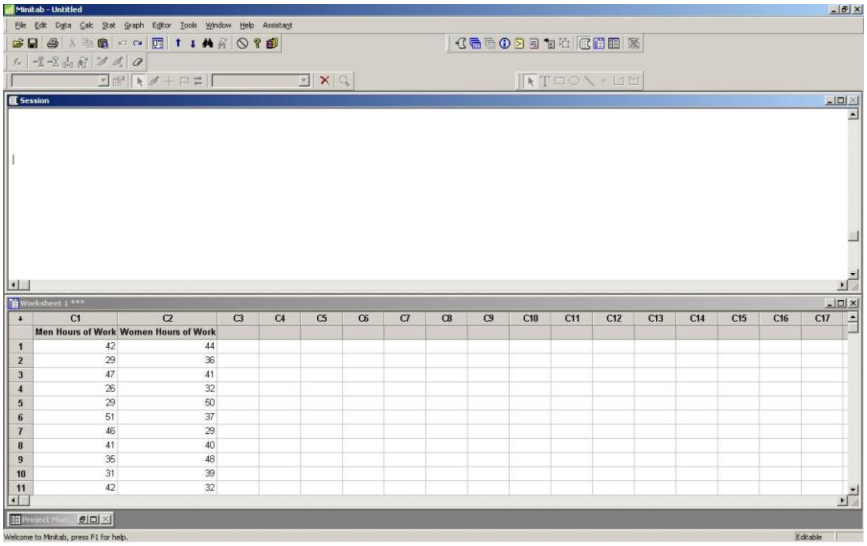
Enter 90 in the **Confidence Level** field. Make sure the **Alternative** field has the **not equal** option selected. Click **OK**. Click **OK** on the **2-Sample t (Test and Confidence Interval)** window.



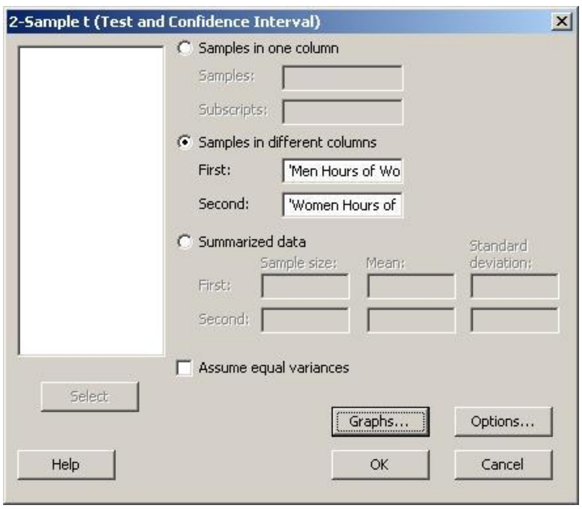
The results will be displayed in the Session window. The 90% confidence interval for the difference between the means is (-2.53, 3.87).



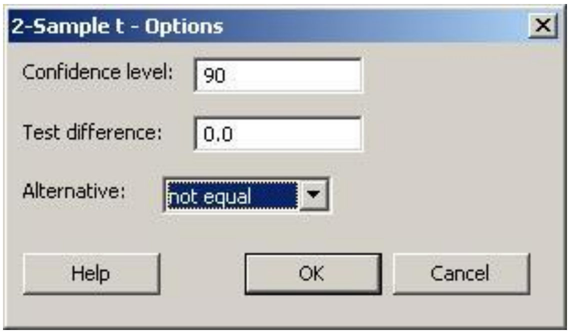
Another way the data can be entered is by entering the data for each group into its own column. Enter the data for the 25 men into column C1 of a Minitab spreadsheet and label the column "Men Hours of Work." Enter the data for the 20 women into column C2 of a Minitab spreadsheet and label the column "Women Hours of Work."



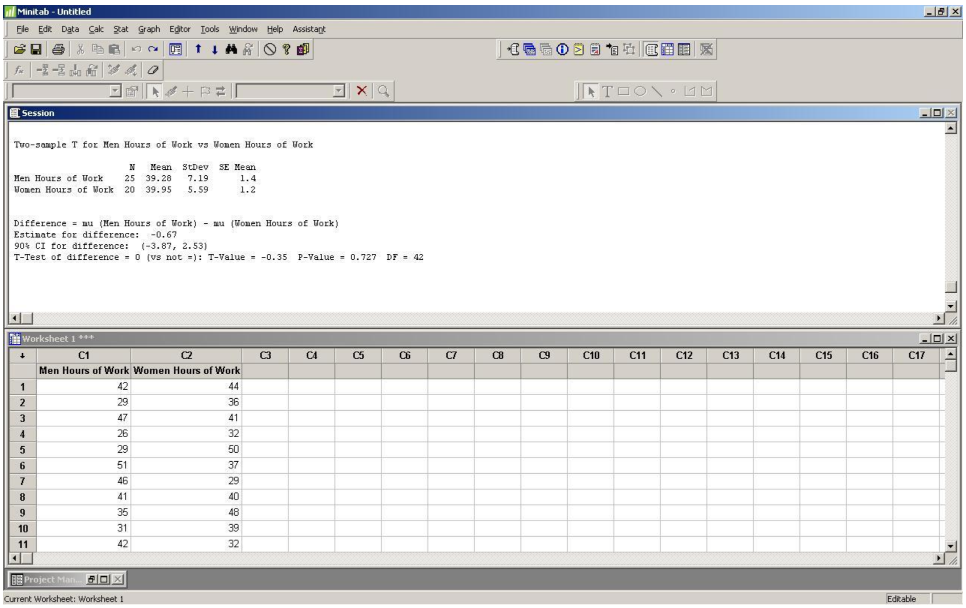
To construct a 90% confidence interval for the mean, click on **Stat → Basic Statistics → 2-Sample t.** Select the circle next to **Samples in different columns** and select the data in column C1 for the **First** sample and the data in column C2 for the **Second** sample. If you assume that the two populations have equal variances, select the **Assume equal variances** checkbox. If you do not have this assumption, do not select the checkbox. Select the **Options** button.



Enter 90 in the **Confidence Level** field. Make sure the **Alternative** field has the **not equal** option selected. Click **OK**. Click **OK** on the **2-Sample t (Test and Confidence Interval)** window.

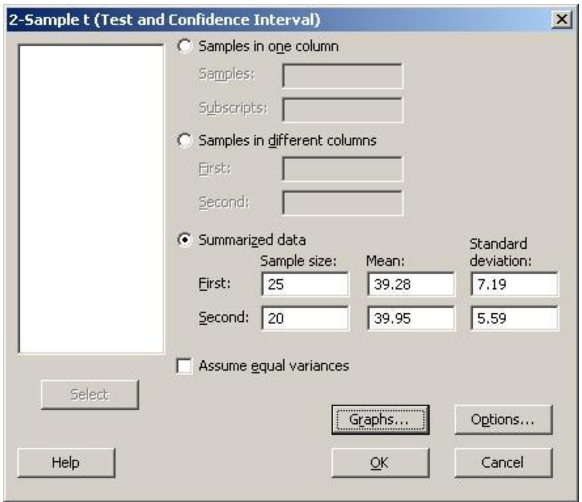


The results will be displayed in the Session window. The 90% confidence interval for the difference between the means is (-3.87, 2.53).

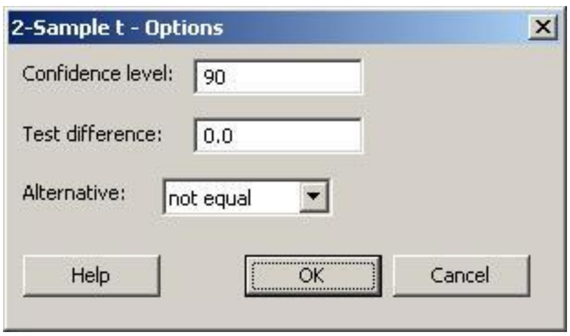


The final way to enter the data is to enter the summary statistics for each group instead of the raw data.

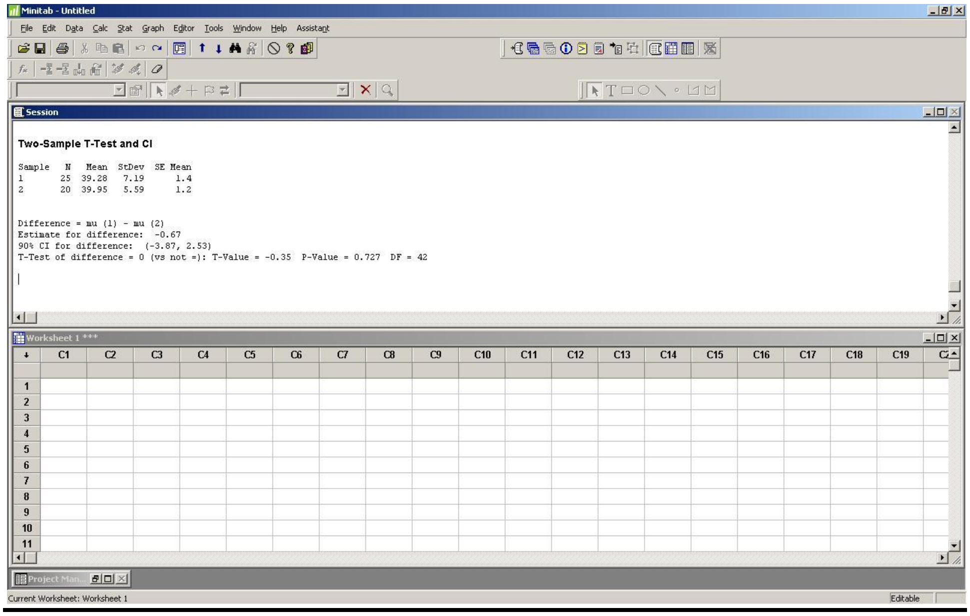
To construct a 90% confidence interval for the mean, click on **Stat → Basic Statistics → 2-Sample t.** Select the circle next to **Summarized data**. For the **First** sample, enter the information for the sample of men. Enter a **Sample size** of 25, a **Mean** of 39.28, and a **Standard deviation** of 7.19. For the **Second** sample, enter the information for the sample of women. Enter a **Sample size** of 20, a **Mean** of 39.95, and a **Standard deviation** of 5.59. If you assume that the two populations have equal variances, select the **Assume equal variances** checkbox. If you do not have this assumption, do not select the checkbox. Select the **Options** button.



Enter 90 in the **Confidence Level** field. Make sure the **Alternative** field has the **not equal** option selected. Click **OK**. Click **OK** on the **2-Sample t (Test and Confidence Interval)** window.



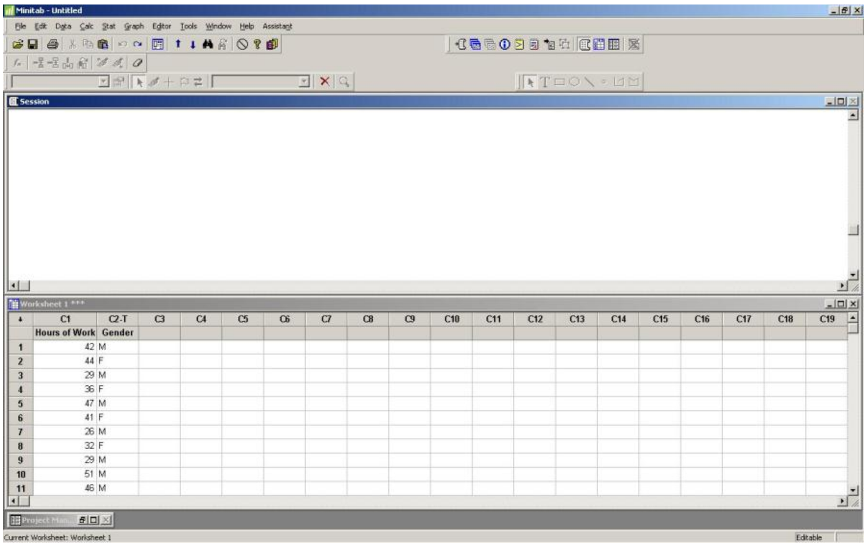
The results will be displayed in the Session window. The 90% confidence interval for the difference between the means is (-3.87, 2.53).



Hypothesis Testing for Independent Samples

Refer to above example for 25 men and 20 women sampled at a corporation. Test if the mean hours of work per week are the same for men and women at this corporation.

Again, there are 3 ways to enter this data into Minitab to be analyzed. The first way is to enter the data into 1 column with an indicator in the second column as to the group each data point belongs. Assume that an indicator of "M" is for a data value of a Male, and "F" is an indicator for a female. Enter the data values in column C1 and label the column "Hours of Work." Enter the indicators in column C2 and label the column "Gender."

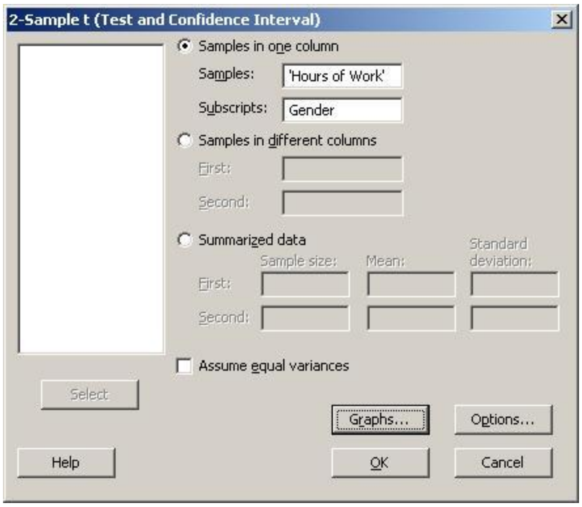


To test the hypothesis

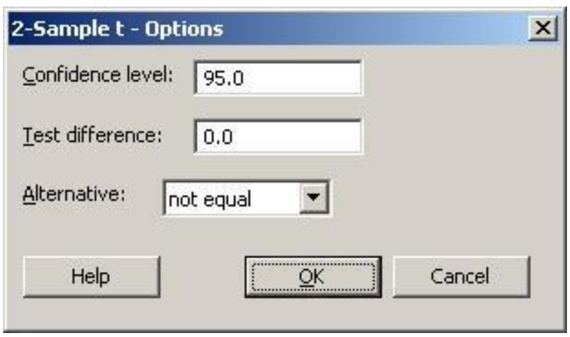
$$H_O : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

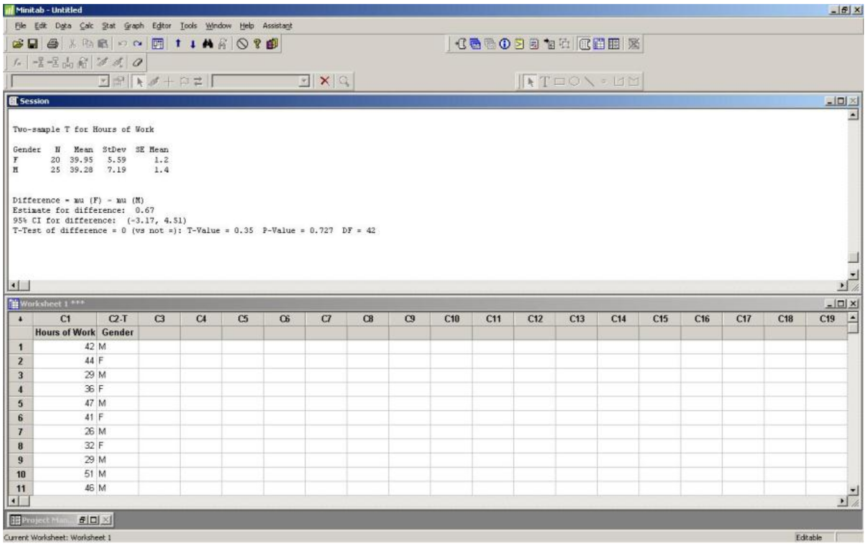
click on **Stat → Basic Statistics → 2-Sample t.** Select the circle next to **Samples in one column.** Select the data in column C1 for the **Samples** field, and the data in column C2 for the **Subscripts** field. If you assume that the two populations have equal variances, select the **Assume equal variances** checkbox. If you do not have this assumption, do not select the checkbox. Select the **Options** button.



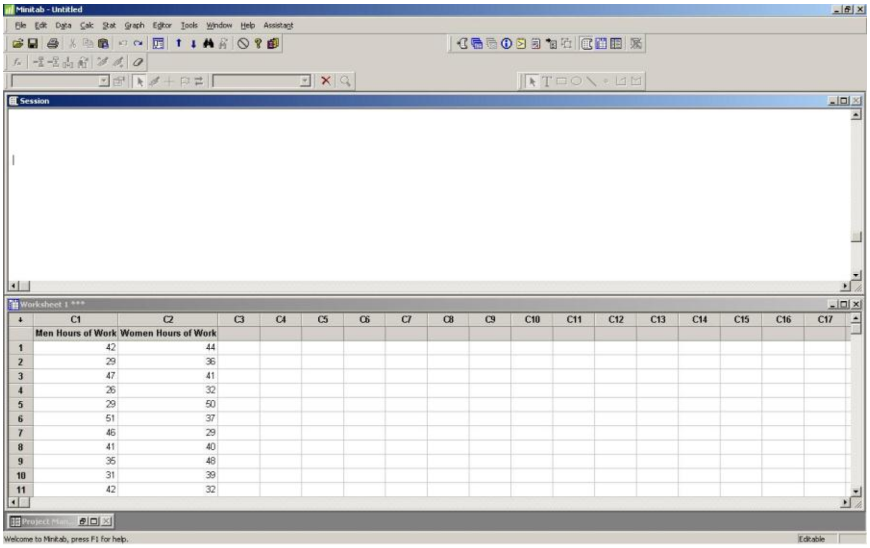
The **Confidence Level** field will be defaulted to 95.0, for a 95% confidence interval. This will not need to be edited this unless you would also like to perform a different confidence interval at this time. This will not influence the results of the hypothesis test. Enter a **Test difference** of 0. Since the alternative hypothesis is "not equal to" for this example, make sure the **Alternative** field has the **not equal** option selected. Click **OK**. Click **OK** on the **2-Sample t (Test and Confidence Interval)** window.



The results will be displayed in the Session window. In the test of "mu (F) – mu (M) = 0 vs not = 0", *T* is 0.35, and the *p*-value is 0.727. Therefore, for any level of $\alpha < 0.727$, the correct decision would be to not reject the null hypothesis.



Another way the data can be entered is by entering the data for each group into its own column. Enter the data for the 25 men into column C1 of a Minitab spreadsheet and label the column "Men Hours of Work." Enter the data for the 20 women into column C2 of a Minitab spreadsheet and label the column "Women Hours of Work."

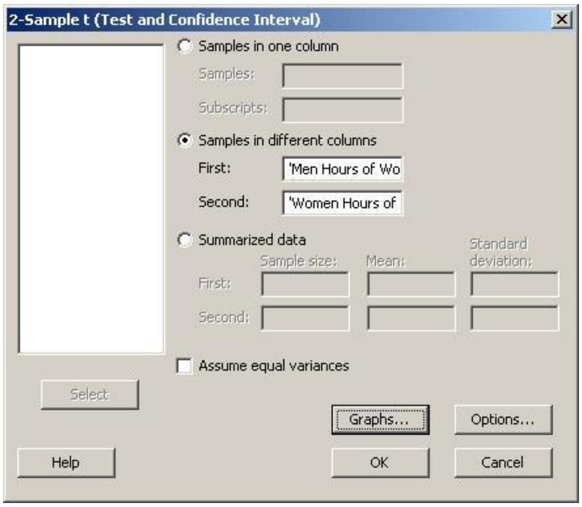


To test the hypothesis

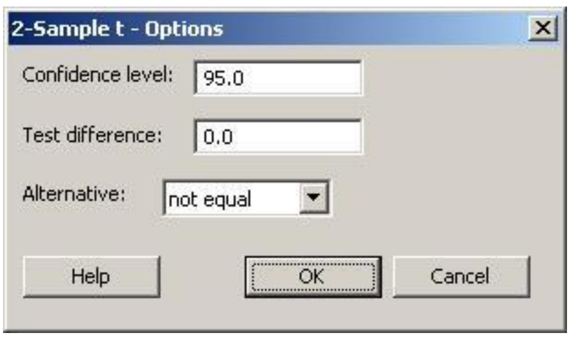
$$H_O : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

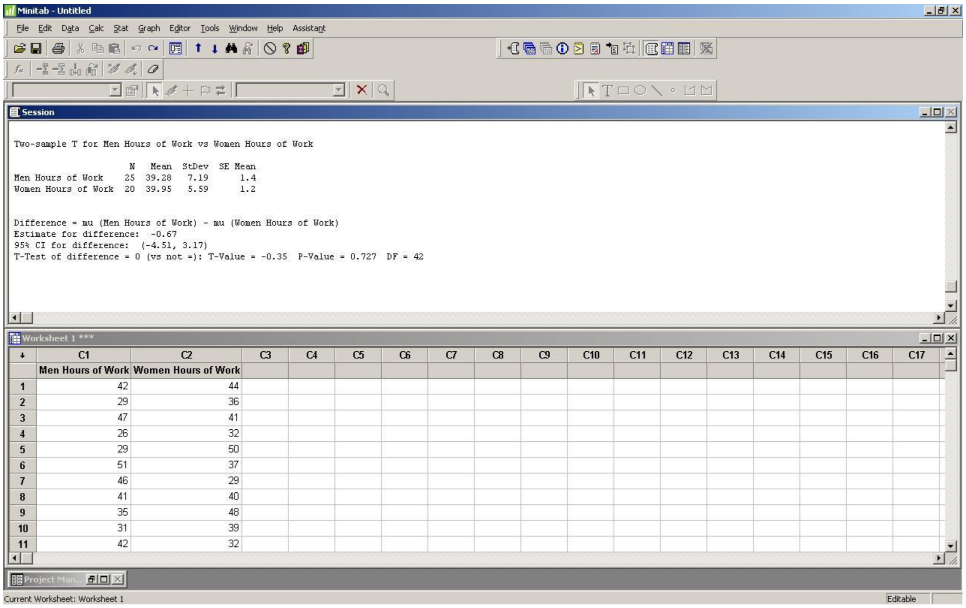
click on **Stat → Basic Statistics → 2-Sample t.** Select the circle next to **Samples in different columns** and select the data in column C1 for the **First** sample and the data in column C2 for the **Second** sample. If you assume that the two populations have equal variances, select the **Assume equal variances** checkbox. If you do not have this assumption, do not select the checkbox. Select the **Options** button.



The **Confidence Level** field will be defaulted to 95.0, for a 95% confidence interval. This will not need to be edited this unless you would also like to perform a different confidence interval at this time. This will not influence the results of the hypothesis test. Enter a **Test difference** of 0. Since the alternative hypothesis is "not equal to" for this example, make sure the **Alternative** field has the **not equal** option selected. Click **OK**. Click **OK** on the **2-Sample t (Test and Confidence Interval)** window.



The results will be displayed in the Session window. In the test of "mu (F) – mu (M) = 0 vs not = 0", *T* is 0.34, and the *p*-value is 0.727. Therefore, for any level of $\alpha < 0.727$, the correct decision would be to not reject the null hypothesis.



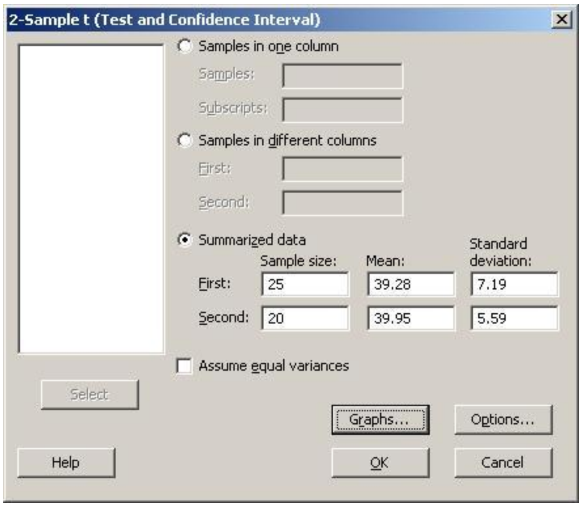
The final way to enter the data is to enter the summary statistics for each group instead of the raw data.

To test the hypothesis

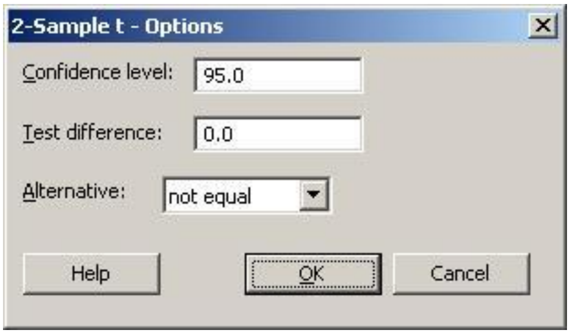
$$H_O : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

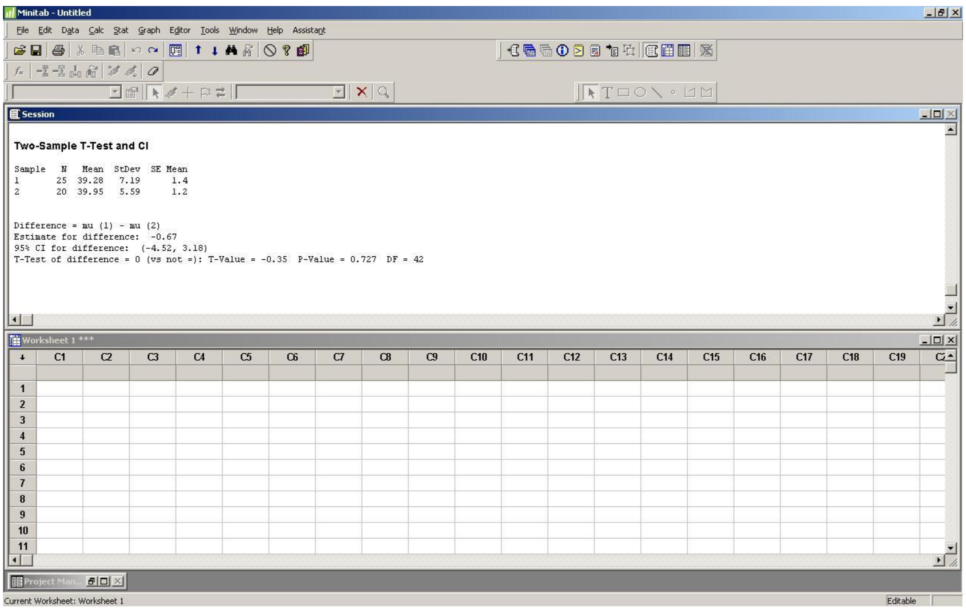
click on **Stat → Basic Statistics → 2-Sample t.** Select the circle next to **Summarized data**. For the **First** sample, enter the information for the sample of men. Enter a **Sample size** of 25, a **Mean** of 39.28, and a **Standard deviation** of 7.19. For the **Second** sample, enter the information for the sample of women. Enter a **Sample size** of 20, a **Mean** of 39.95, and a **Standard deviation** of 5.59. If you assume that the two populations have equal variances, select the **Assume equal variances** checkbox. If you do not have this assumption, do not select the checkbox. Select the **Options** button.



The **Confidence Level** field will be defaulted to 95.0, for a 95% confidence interval. This will not need to be edited this unless you would also like to perform a different confidence interval at this time. This will not influence the results of the hypothesis test. Enter a **Test difference** of 0. Since the alternative hypothesis is "not equal to" for this example, make sure the **Alternative** field has the **not equal** option selected. Click **OK**. Click **OK** on the **2-Sample t (Test and Confidence Interval)** window.



The results will be displayed in the Session window. In the test of "mu (F) – mu (M) = 0 vs not = 0", T is 0.35, and the p-value is 0.727. Therefore, for any level of $\alpha < 0.727$, the correct decision would be to not reject the null hypothesis.



Inferences Concerning Paired Samples

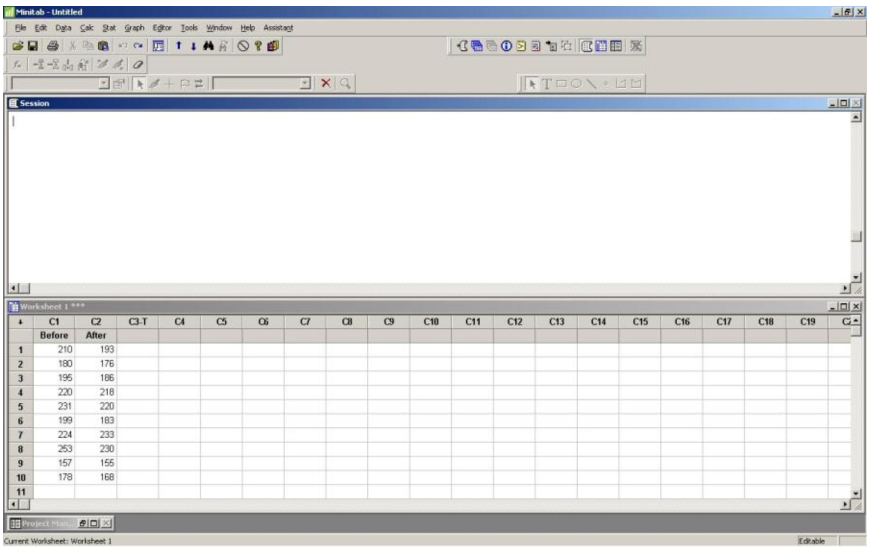
It is important to note that Minitab does not offer a Paired *Z*-test. Instead, the Paired *t*-test is used to construct any inferences for μ_d.

Constructing a Confidence Interval for Paired Samples

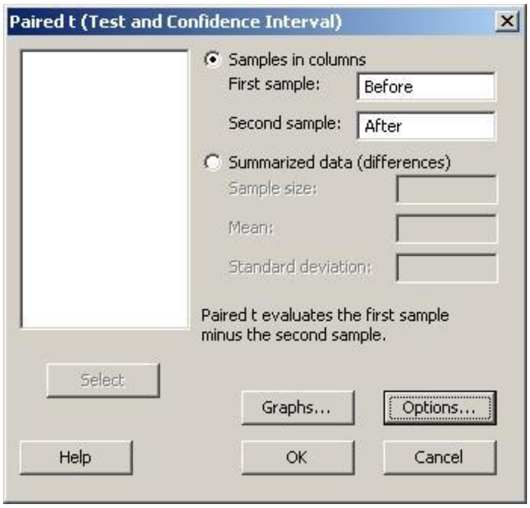
A researcher wanted to find the effect of a certain exercise program on weight loss. Ten adults were selected, and assigned to perform the exercise program for two months. The data shows the weight of the ten adults before and after the exercise program.

Before	210	180	195	220	231	199	224	253	157	178
After	193	176	186	218	220	183	233	230	155	168

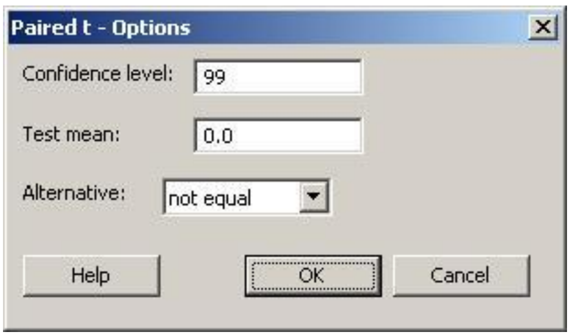
There are two ways the data can be entered into Minitab to be analyzed. The first is to enter the raw data into separate columns for each treatment. Enter the Before data into column C1 of a Minitab spreadsheet. Enter the After data into column C2 of a Minitab spreadsheet. Make sure that each pair of data is in the same row of the list of data.



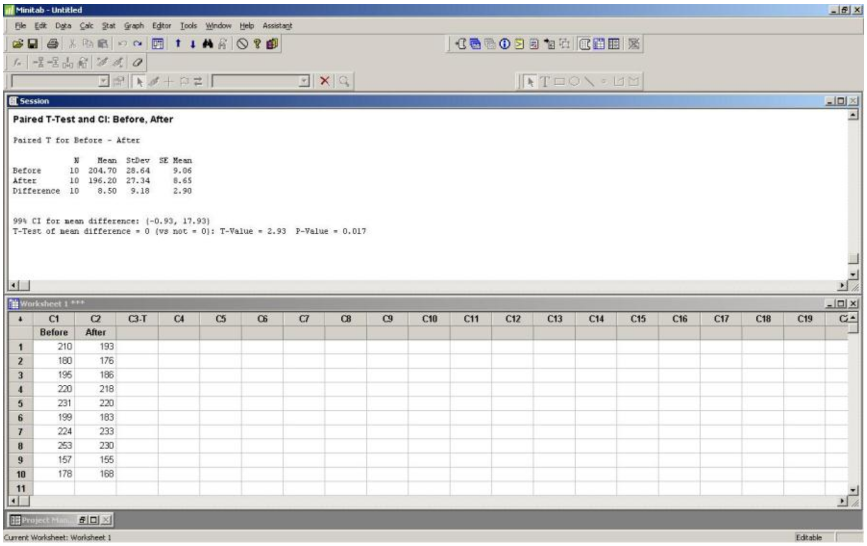
To construct a 99% confidence interval for the mean difference between the weights, click on **Stat → Basic Statistics → Paired t.** Select the circle next to **Samples in columns** and select the Before data in column C1 for the **First sample** and the After data in column C2 for the **Second sample**. Select the **Options** button.



Enter 99 in the **Confidence Level** field. Make sure the **Alternative** field has the **not equal** option selected. Click **OK**. Click **OK** on the **Paired t (Test and Confidence Interval)** window.

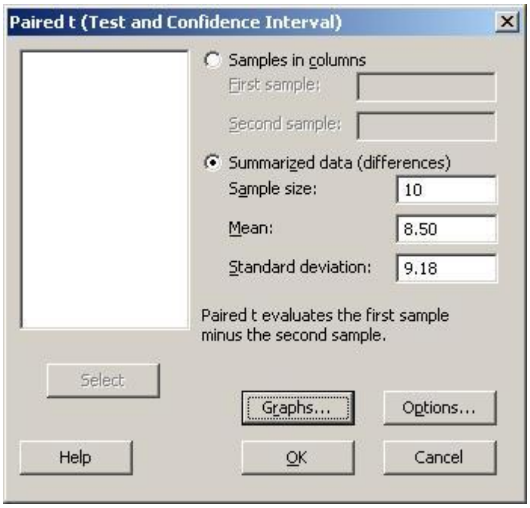


The results will be displayed in the Session window. The 99% confidence interval for the mean difference is (-0.93, 17.93).

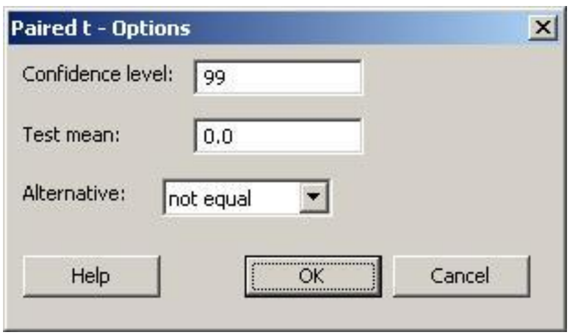


Another way to enter the data is to enter the summary statistics for each group instead of the raw data.

To construct a 90% confidence interval for the mean, click on **Stat → Basic Statistics → 2-Sample t.** Select the circle next to **Summarized data (differences)**. Enter a **Sample Size** of 10, a **Mean** of 8.50, and a **Standard deviation** of 9.18. Select the **Options** button.



Enter 99 in the **Confidence Level** field. Make sure the **Alternative** field has the **not equal** option selected. Click **OK**. Click **OK** on the **Paired t (Test and Confidence Interval)** window.

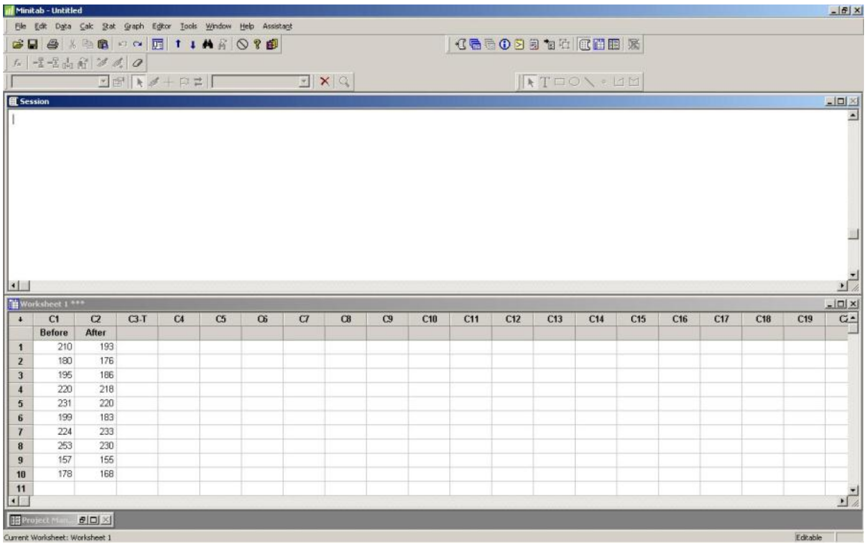


The results will be displayed in the Session window. The 99% confidence interval for the mean difference is (-0.93, 17.93).

Hypothesis Testing for Paired Samples

Refer to above example for 10 adults assigned to the exercise program for two months. Test if the mean weight loss exceeds 5 pounds after two months.

Again, there are two ways to enter this data into Minitab to be analyzed. When using the raw data, enter the Before data into column C1 of a Minitab spreadsheet. Enter the After data into column C2 of a Minitab spreadsheet. Make sure that each pair of data is in the same row of the list of data.

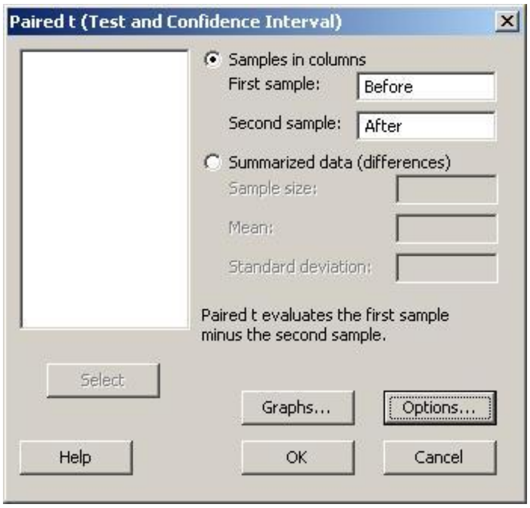


To test the hypothesis

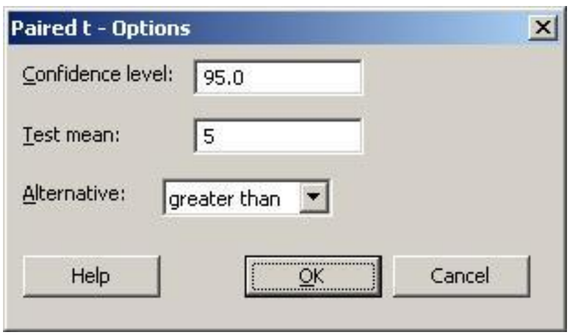
$$H_O : \mu_d = 5$$

$$H_1 : \mu_d > 5$$

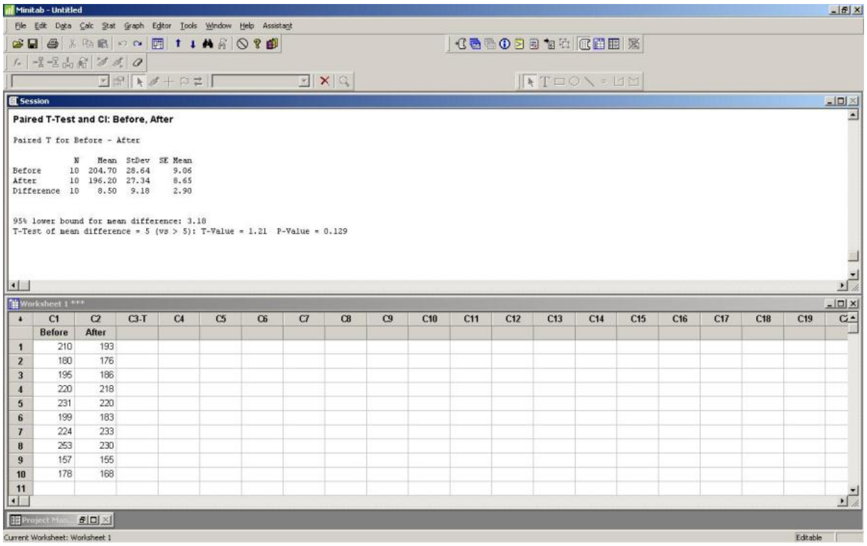
click on **Stat → Basic Statistics → Paired t.** Select the circle next to **Samples in columns** and select the Before data in column C1 for the **First sample** and the After data in column C2 for the **Second sample**. Select the **Options** button.



The **Confidence Level** field will be defaulted to 95.0, for a 95% confidence interval. This will not need to be edited this unless you would also like to perform a different confidence interval at this time. This will not influence the results of the hypothesis test. Enter a **Test mean** of 5. Since the alternative hypothesis is "greater than" for this example, make sure the **Alternative** field has the **greater than** option selected. Click **OK**. Click **OK** on the **Paired t (Test and Confidence Interval)** window.



The results will be displayed in the Session window. In the test of "mean difference = 5 vs > 5", T is 1.21, and the p-value is 0.129. Therefore, for any level of $\alpha < 0.129$, the correct decision would be to not reject the null hypothesis.



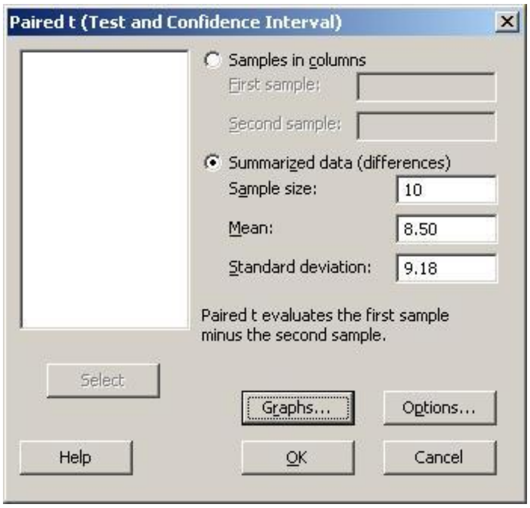
Another way to enter the data is to enter the summary statistics for each group instead of the raw data.

To test the hypothesis

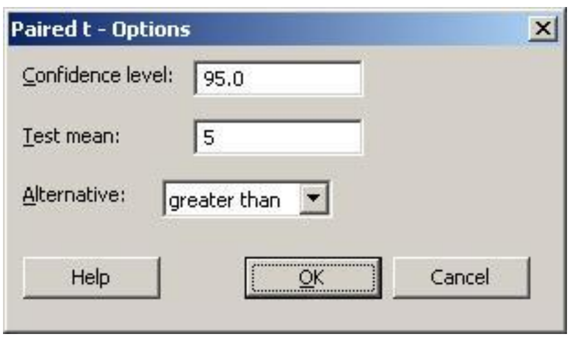
$$H_O : \mu_d = 5$$

$$H_1 : \mu_d > 5$$

click on **Stat** → **Basic Statistics** → **2-Sample t.** Select the circle next to **Summarized data (differences)**. Enter a **Sample Size** of 10, a **Mean** of 8.50, and a **Standard Deviation** of 9.18. Select the **Options** button.



The **Confidence Level** field will be defaulted to 95.0, for a 95% confidence interval. This will not need to be edited this unless you would also like to perform a different confidence interval at this time. This will not influence the results of the hypothesis test. Enter a **Test mean** of 5. Since the alternative hypothesis is "greater than" for this example, make sure the **Alternative** field has the **greater than** option selected. Click **OK**. Click **OK** on the **Paired t (Test and Confidence Interval)** window.



The results will be displayed in the Session window. In the test of "mean difference = 5 vs > 5", *T* is 1.21, and the *p*-value is 0.129. Therefore, for any level of $\alpha < 0.129$, the correct decision would be to not reject the null hypothesis.

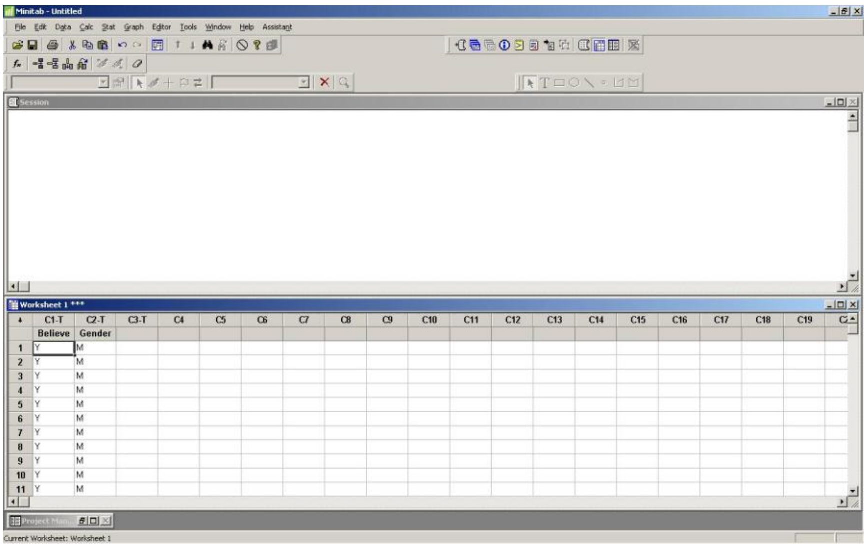
Inferences Concerning the Difference Between Two Proportions

Constructing a Confidence Interval for The Difference Between Two Proportions

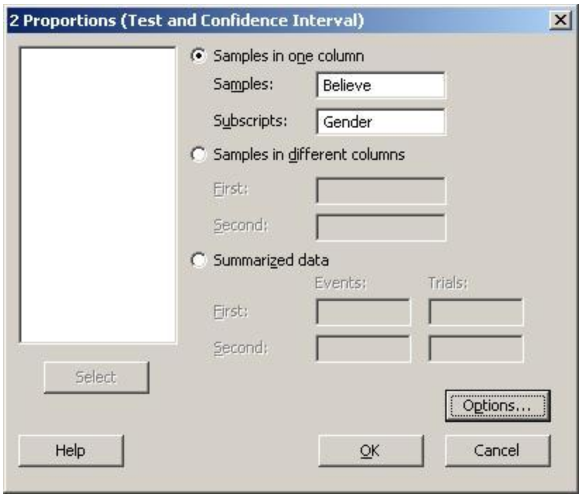
In a random sample of 35 adult males, it was found that 18 believe in ghosts. In a random sample of 30 adult women, it was found that 21 believe in ghosts. A researcher is interested in the difference between the proportion of men and women who believe in ghosts.

There are 3 ways to enter this data into Minitab to be analyzed. The first way is to enter the data into 1 column with an indicator in the second column as to the group each data point belongs. Assume that a "Y" refers to the person responding that they do believe in ghosts, while a "N" refers to the person responding that they do not believe in ghosts. Assume that an indicator of "M" is for a data value of a Male, and "F" is an indicator for a female. Enter the data values in column C1 and label the column "Believe in Ghosts." Enter the indicators in column C2 and label the column "Gender."

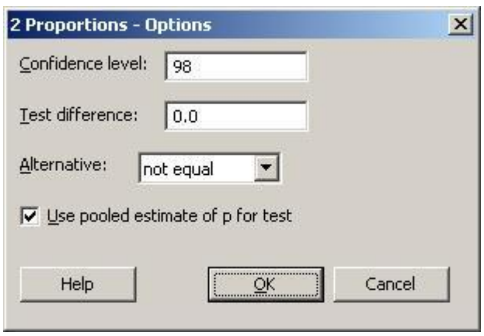
The data will be 75 rows, with 18 "Y" and "M" (males who believe), 17 "N" and "M" (males who do not believe), 21 "Y" and "F" (females who believe), and 9 "N" and "F" (females who do not believe).



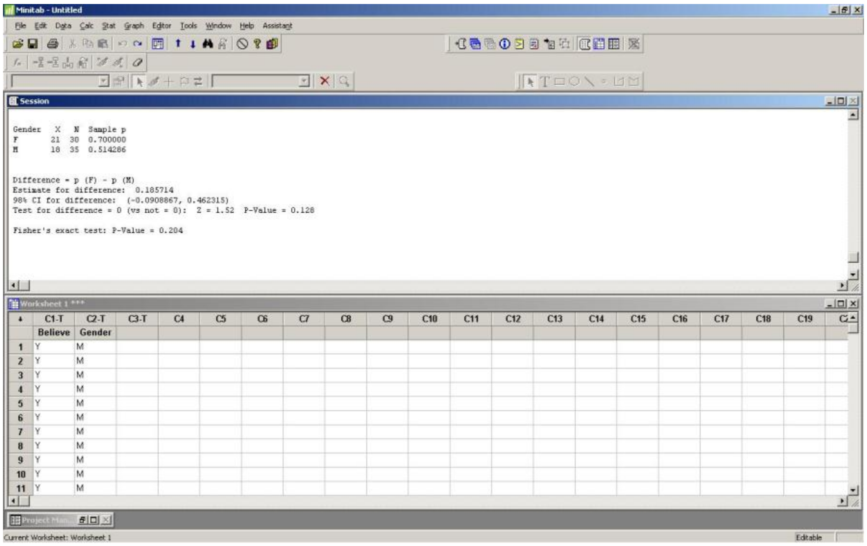
To construct a 98% confidence interval for the mean, click on **Stat → Basic Statistics → 2 Proportions.** Select the circle next to **Samples in one column** and select the data in column C1 for the **Samples** and the data in column C2 for the **Subscripts**. Select the **Options** button.



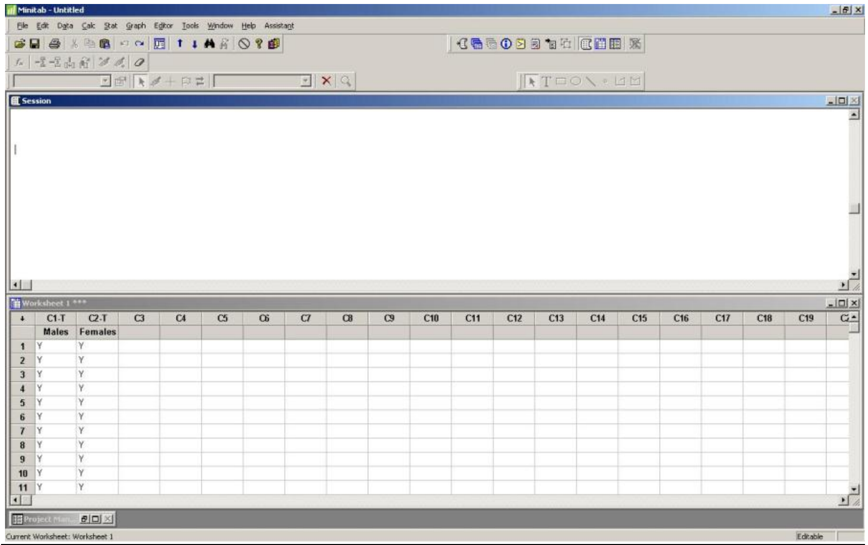
Enter 98 in the **Confidence Level** field. Make sure the **Alternative** field has the **not equal** option selected. Select the checkbox to **Use pooled estimate of p for test**. Click **OK**. Click **OK** on the **2 Proportions (Test and Confidence Interval)** window.



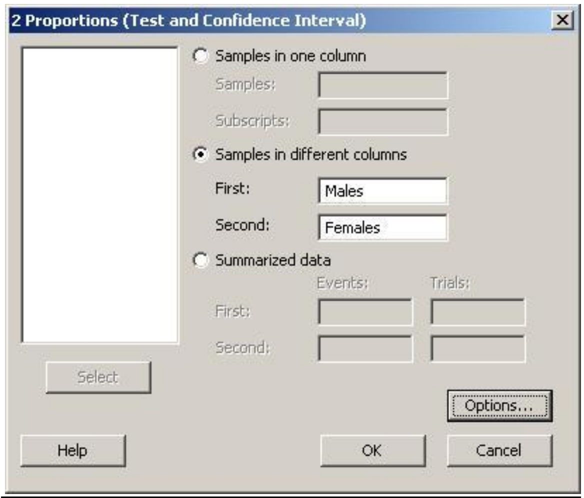
The results will be displayed in the Session window. The 98% confidence interval for the difference between the proportions is (-0.0908867, 0.462315).



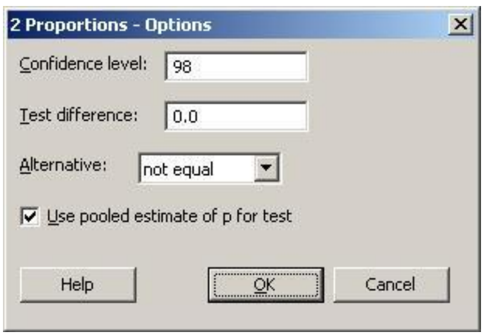
Another way the data can be entered is by entering the data for each group into its own column. Enter the data for the 35 men into column C1 of a Minitab spreadsheet and label the column "Males". The column should contain 18 "Y" and 17 "N." Enter the data for the 30 women into column C2 of a Minitab spreadsheet and label the column "Females". The column should contain 21 "Y" and 9 "N."



To construct a 98% confidence interval for the difference between the proportions, click on **Stat → Basic Statistics → 2 Proportions.** Select the circle next to **Samples in different columns** and select the data in column C1 for the **First** sample and the data in column C2 for the **Second** sample. Select the **Options** button.



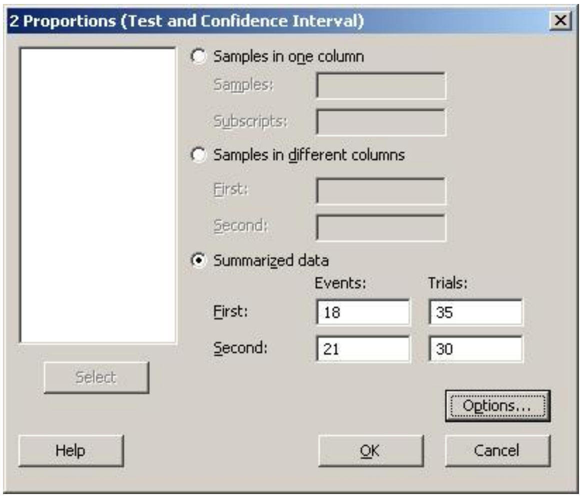
Enter 98 in the **Confidence Level** field. Make sure the **Alternative** field has the **not equal** option selected. Select the checkbox to **Use pooled estimate of p for test**. Click **OK**. Click **OK** on the **2 Proportions (Test and Confidence Interval)** window.



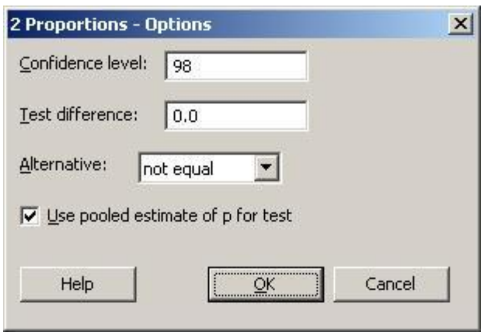
The results will be displayed in the Session window. The 98% confidence interval for the difference between the proportions is (-0.0908867, 0.462315).

The final way to enter the data is to enter the summary statistics for each group instead of the raw data.

To construct a 98% confidence interval for the difference between the proportions, click on **Stat → Basic Statistics → 2 Proportions.** Select the circle next to **Summarized data.** For the **first** sample, enter 18 **Events** and 35 **Trials.** For the **second** sample, enter 21 **Events** and 30 **Trials**. Select the **Options** button.



Enter 98 in the **Confidence Level** field. Make sure the **Alternative** field has the **not equal** option selected. Select the checkbox to **Use pooled estimate of p for test**. Click **OK**. Click **OK** on the **2 Proportions (Test and Confidence Interval)** window.

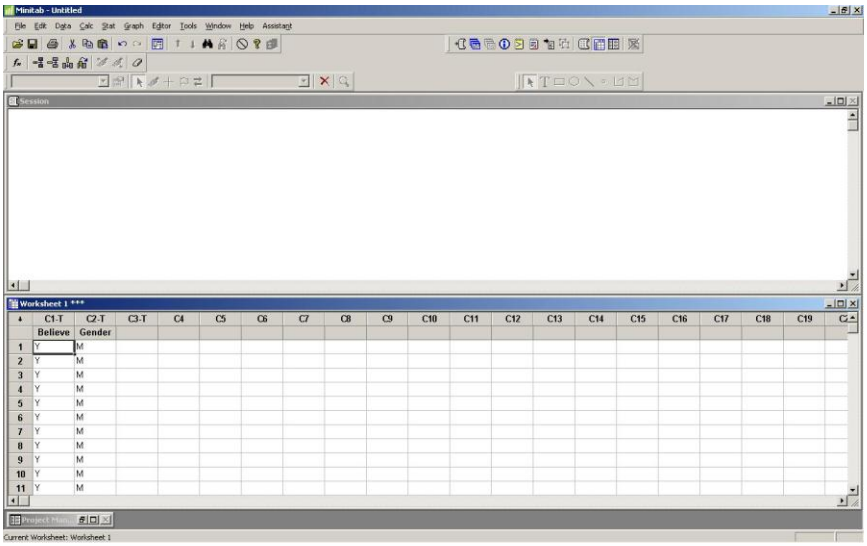


The results will be displayed in the Session window. The 98% confidence interval for the mean difference is (-0.0908867, 0.462315).

Hypothesis Testing for the Difference Between Two Proportions

Using the above data concerning the males and females belief in ghosts, a researcher wants to test that the proportion of women who believe in ghosts is greater than the proportion of men. There are three ways to enter this data into Minitab to be analyzed. The first way is to enter the data into 1 column with an indicator in the second column as to the group each data point belongs. Assume that a “Y” refers to the person responding that they do believe in ghosts, while a “N” refers to the person responding that they do not believe in ghosts. Assume that an indicator of “M” is for a data value of a Male, and “F” is an indicator for a female. Enter the data values in column C1 and label the column “Believe in Ghosts.” Enter the indicators in column C2 and label the column “Gender.”

The data will be 75 rows, with 18 “Y” and “M” (males who believe), 17 “N” and “M” (males who do not believe), 21 “Y” and “F” (females who believe), and 9 “N” and “F” (females who do not believe).

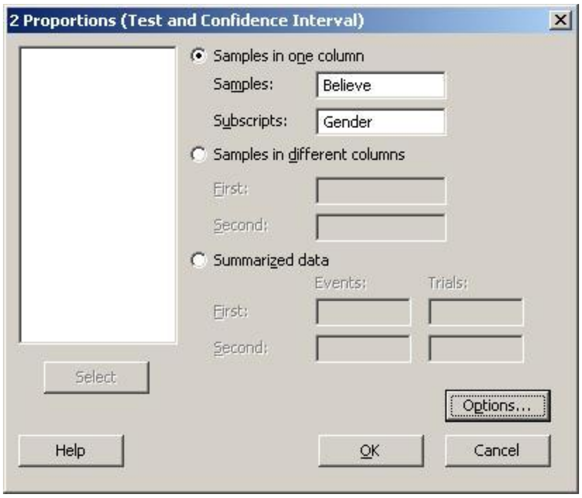


To test the hypothesis

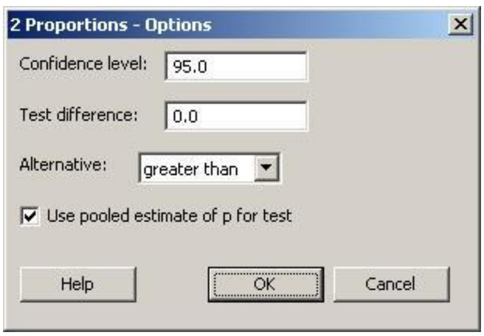
$$H_O : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 > 0$$

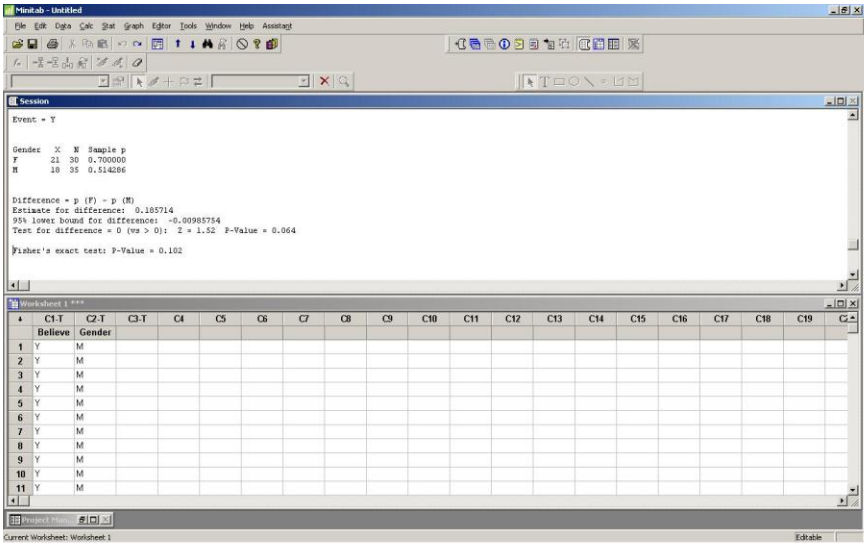
click on **Stat → Basic Statistics → 2 Proportions.** Select the circle next to **Samples in one column** and select the data in column C1 for the **Samples** and the data in column C2 for the **Subscripts**. Select the **Options** button.



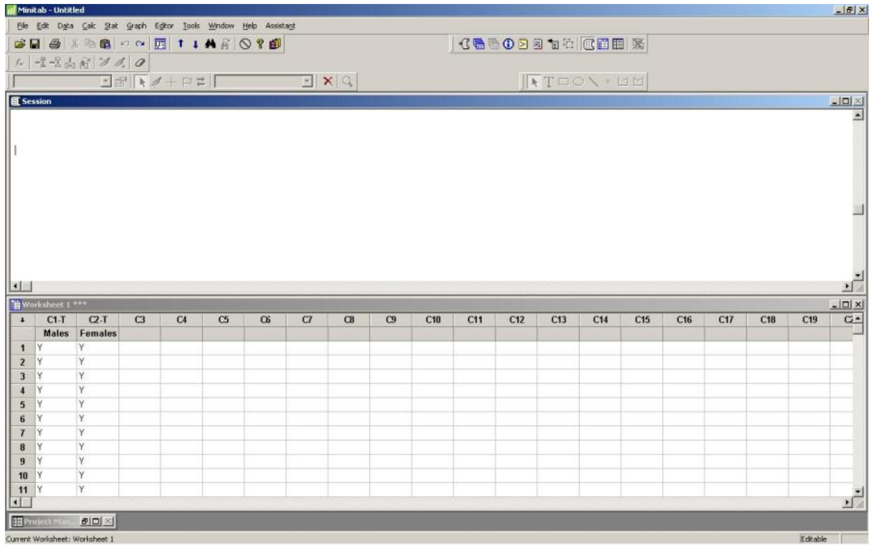
The **Confidence Level** field will be defaulted to 95.0, for a 95% confidence interval. This will not need to be edited this unless you would also like to perform a different confidence interval at this time. This will not influence the results of the hypothesis test. Enter a **Test difference** of 0. Since the alternative hypothesis is "less than" for this example, make sure the **Alternative** field has the **greater than** option selected. Select the checkbox to **Use pooled estimate of p for test**. Click **OK**. Click **OK** on the **2 Proportions (Test and Confidence Interval)** window.



The results will be displayed in the Session window. In the test of “p(Females) – p(Males) =0 vs > 0”, Z is 1.52, and the p-value is 0.064.   Therefore, for any level of $\alpha > 0.064$, the correct decision would be to reject the null hypothesis.



Another way the data can be entered is by entering the data for each group into its own column. Enter the data for the 35 men into column C1 of a Minitab spreadsheet and label the column "Males." The column should contain 18 "Y" and 17 "N." Enter the data for the 30 women into column C2 of a Minitab spreadsheet and label the column "Females." The column should contain 21 "Y" and 9 "N."

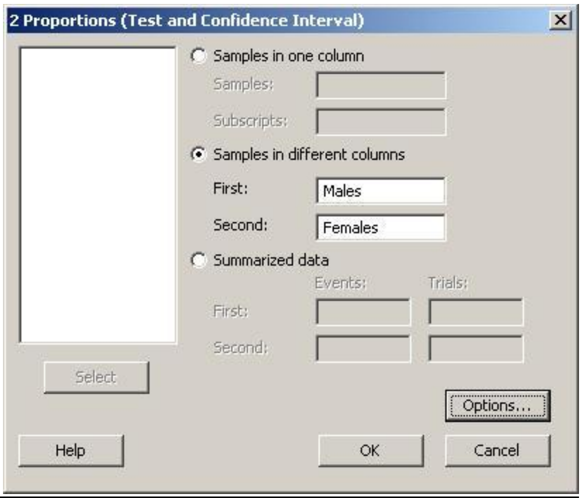


To test the hypothesis

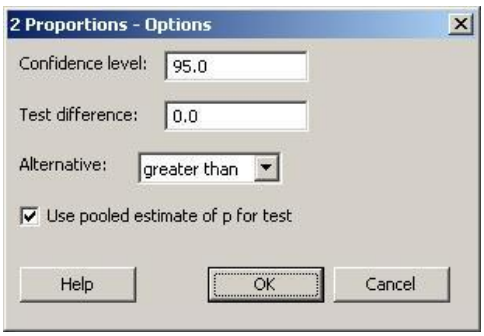
$$H_O : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 > 0$$

click on **Stat → Basic Statistics → 2 Proportions.** Select the circle next to **Samples in different columns** and select the data in column C1 for the **First** sample and the data in column C2 for the **Second** sample. Select the **Options** button.



The **Confidence Level** field will be defaulted to 95.0, for a 95% confidence interval. This will not need to be edited this unless you would also like to perform a different confidence interval at this time. This will not influence the results of the hypothesis test. Enter a **Test difference** of 0. Since the alternative hypothesis is "greater than" for this example, make sure the **Alternative** field has the **greater than** option selected. Select the checkbox to **Use pooled estimate of p for test**. Click **OK**. Click **OK** on the **2 Proportions (Test and Confidence Interval)** window.



The results will be displayed in the Session window. In the test of "p(Females) – p(Males) =0 vs > 0", *Z* is 1.52, and the *p*-value is 0.064. Therefore, for any level of $\alpha > 0.064$, the correct decision would be to reject the null hypothesis.

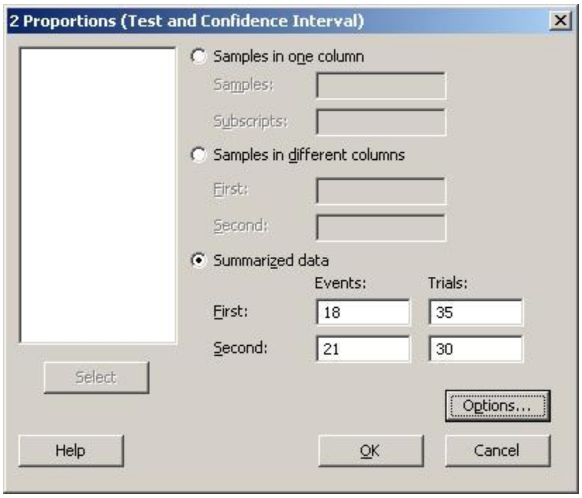
The final way to enter the data is to enter the summary statistics for each group instead of the raw data.

To test the hypothesis

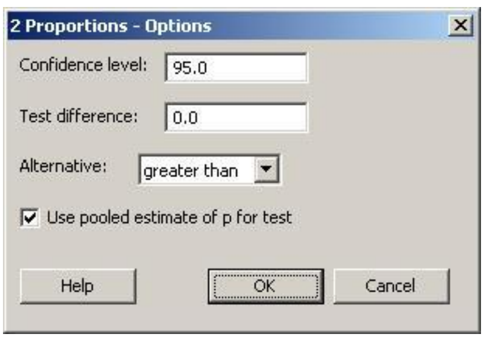
$$H_O : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 > 0$$

click on **Stat → Basic Statistics → 2 Proportions.** Select the circle next to **Summarized data.** For the **first** sample, enter 18 **Events** and 35 **Trials.** For the **second** sample, enter 21 **Events** and 30 **Trials**. Select the **Options** button.



The **Confidence Level** field will be defaulted to 95.0, for a 95% confidence interval. This will not need to be edited this unless you would also like to perform a different confidence interval at this time. This will not influence the results of the hypothesis test. Enter a **Test difference** of 0. Since the alternative hypothesis is "greater than" for this example, make sure the **Alternative** field has the **less than** option selected. Select the checkbox to **Use pooled estimate of p for test**. Click **OK**. Click **OK** on the **2 Proportions (Test and Confidence Interval)** window.



The results will be displayed in the Session window. In the test of "p(Females) – p(Males) =0 vs > 0", *Z* is 1.52, and the *p*-value is 0.064. Therefore, for any level of $\alpha > 0.064$, the correct decision would be to reject the null hypothesis.

Suggested Exercises

Section 10.1

10.9, 10.10, 10.11, 10.14, 10.15

Section 10.2

10.25, 10.26, 10.27, 10.29, 10.32

Section 10.3

10.39, 10.40, 10.41, 10.43, 10.46

Section 10.4

10.52, 10.53, 10.54, 10.55, 10.57

Section 10.5

10.68, 10.70, 10.71, 10.74, 10.75

Supplementary Exercises

10.76, 10.77, 10.78, 10.79, 10.82, 10.83, 10.86, 10.87, 10.90, 10.94, 10.96

Technology Assignments

TA 10.2, TA 10.4, TA 10.5, TA 10.7, TA 10.8