## INTRODUCTION

Minitab 17 gives you the confidence you need to improve quality, with features like an interactive Assistant that guides you through your analysis. Sometimes your biggest challenge is knowing where to begin. Minitab's Assistant removes any doubt by helping you determine the correct statistical analysis to use. Once you've chosen the right tool, the Assistant identifies all the steps you should take to ensure the results of your analysis are accurate and trustworthy.

More detail about the statistical processes described in this card can be found in the Help menu. This tip sheet will provide tips for analyses completed in introductory and intermediate business statistics courses.

MINITAB ASSISTANT
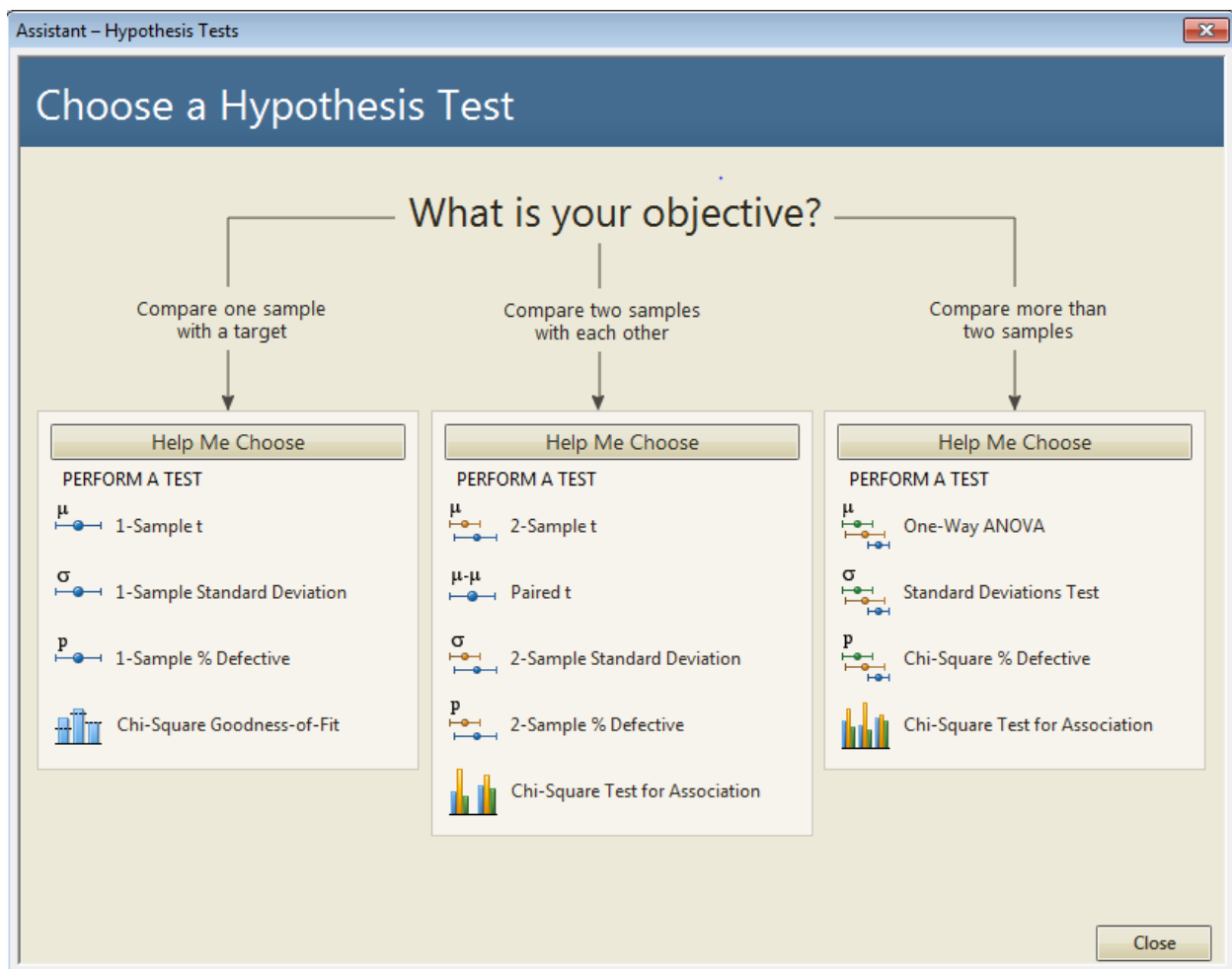
DATA SAMPLING

Use the Sample from Columns tool found in the Random Data option of the Calc pull-down menu to generate a subsample of observations from a dataset. Sample From Columns randomly samples the same rows from one or more columns. You can sample with replacement (select the same row more than once), or without replacement (select each row only once).
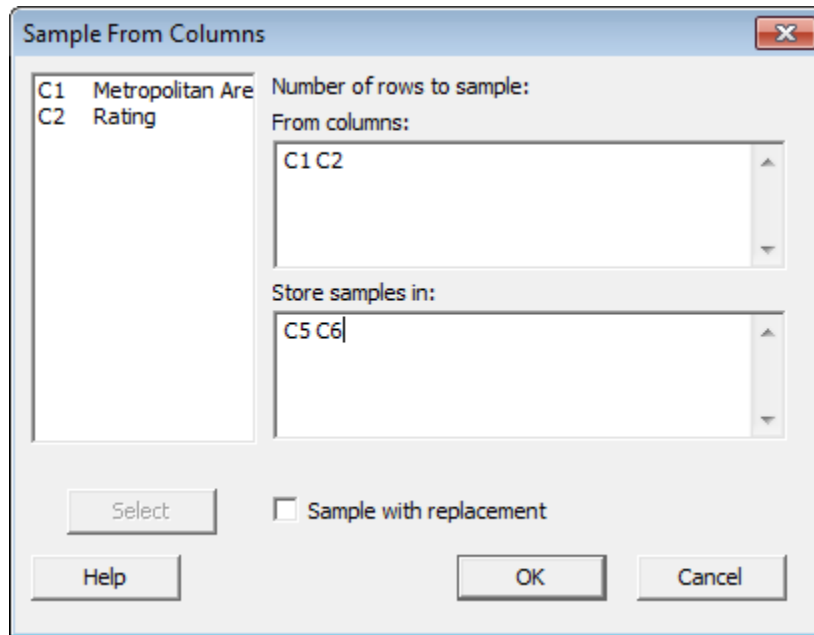
DIALOG BOX ITEMS

Number of rows to sample: Specify the number of rows to randomly select.

From columns: Enter the column(s) you want to sample from. If you sample from several columns at once, they must all have the same length.

Store samples in: Specify the column(s) where you want to store the sampled values. The number of storage columns must be the same as the number of columns sampled from.

Sample with replacement: Check to sample with replacement. Leave unchecked to sample without replacement (sample size must be less than or equal to the length of the columns).
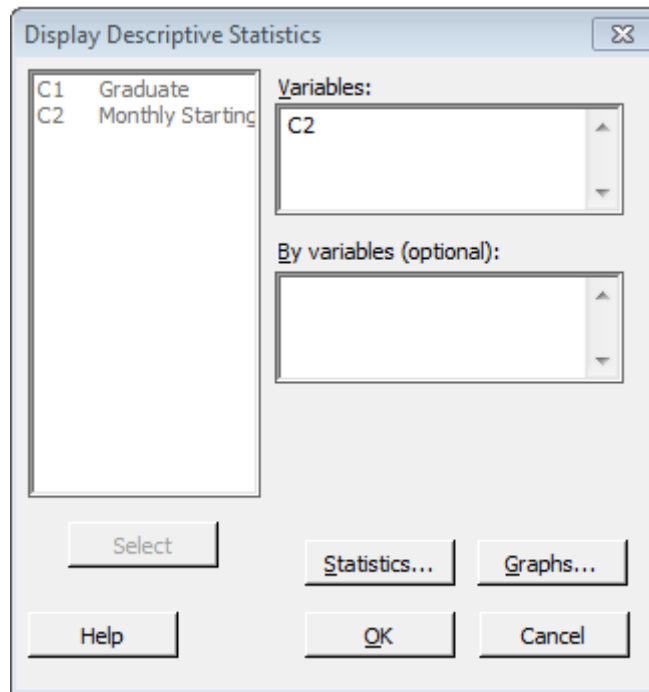
MINITAB BASIC STATISTICS

Use Minitab's basic statistics capabilities for calculating basic statistics and for simple estimation and hypothesis testing with one or two samples. The basic statistics capabilities include procedures for:
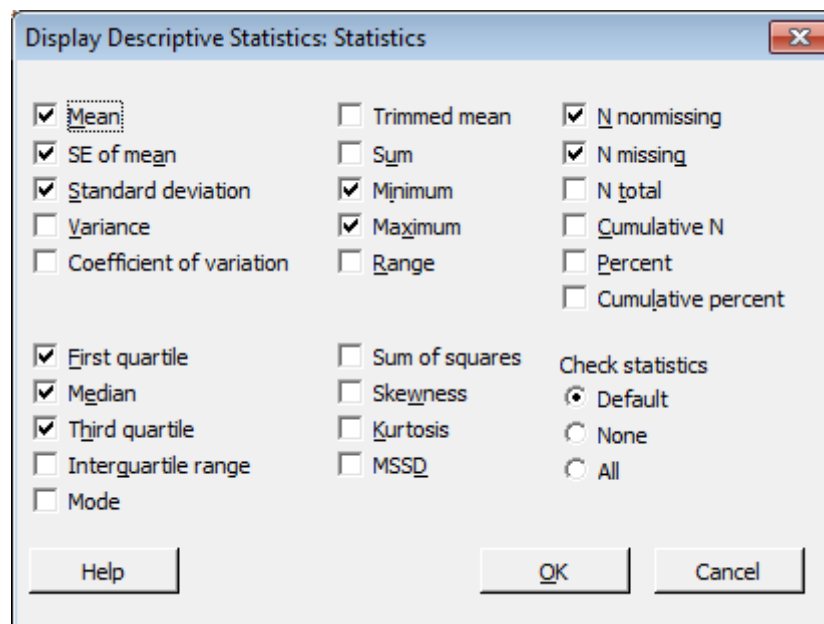
· Calculating or storing descriptive statistics

· Hypothesis tests and confidence intervals of the mean or difference in means

· Hypothesis tests and confidence intervals for a proportion or the difference in proportions

· Hypothesis tests and confidence intervals for one variance , and for the difference between two variances

· Measuring association

· Testing for normality of a distribution

DESCRIPTIVE STATISTICS

The Stat > Basic Statistics > Display Descriptive Statistics command produces descriptive statistics for each column. These statistics can be displayed in the Session window and optionally in a graph.

Display Descriptive Statistics

C1 Graduate
C2 Monthly Starting

Variables:

C2

By variables (optional):

Select

Statistics...   Graphs...

Help   OK   Cancel

DIALOG BOX ITEMS

Display Descriptive Statistics: Statistics

☑ Mean   ☐ Trimmed mean   ☑ N nonmissing
☑ SE of mean   ☐ Sum   ☑ N missing
☑ Standard deviation ☑ Minimum   ☐ N total
☐ Variance   ☑ Maximum   ☐ Cumulative N
☐ Coefficient of variation ☐ Range   ☐ Percent
                                            ☐ Cumulative percent

☑ First quartile   ☐ Sum of squares   Check statistics
☑ Median   ☐ Skewness   ⦿ Default
☑ Third quartile   ☐ Kurtosis   ○ None
☐ Interquartile range ☐ MSSD   ○ All
☐ Mode

Help   OK   Cancel

The Display Descriptive Statistics: Statistics dialog box allows you to display the statistics you wish to display. The **Default** option automatically checks the default set of statistics previously specified in **Tools > Options > Individual Commands > Display Descriptive Statistics**.

```
Session

Results for: Worksheet 4

Descriptive Statistics: Monthly Starting Salary ($)

Variable                  N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3
Monthly Starting Salary  12   0  3940.0     47.8  165.7   3710.0  3857.5  3905.0  4025.0

Variable                 Maximum
Monthly Starting Salary   4325.0
```

Heading definitions for the descriptive statistics are:

|			
N	number of observations	Minimum	minimum data value
N*	number of missing values	Q1	first quartile
SE	Mean standard error of the mean	Q3	third quartile
StDev	standard deviation	Maximum	maximum data value

ONE-SAMPLE Z FOR THE MEAN

The Stat > Basic Statistics > 1-Sample Z for the Mean command computes a confidence interval or performs a hypothesis test of the mean when the population standard deviation, σ, is known. This procedure is based upon the normal distribution, so for small samples, this procedure works best if your data were drawn from a normal distribution or one that is close to normal. A common rule of thumb is to consider samples of size 30 or higher to be large samples. Many analysts choose the t-procedure over the Z-procedure whenever σ is unknown.

For a two-tailed one-sample Z-test, the hypotheses are: $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ where μ is the population mean and μ_0 is the hypothesized population mean.

Choose **One or more samples, each in a column:** if you have entered raw data in columns. Enter the columns containing the sample data. Enter the population known (σ) standard deviation. Check **Perform hypothesis test:** to perform the hypothesis test**.** Enter the **Hypothesized mean:** μ_0. Choose the Options… command.
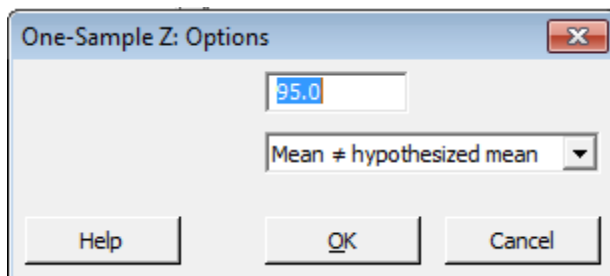
In the **One-Sample Z: Options** dialog box specify the confidence level for the confidence interval, or define the alternative hypothesis. Choose *Mean < hypothesized mean* (lower-tailed), *Mean ≠ hypothesized mean* (two-tailed), or *Mean > hypothesized mean* (upper-tailed). If you choose a lower-tailed or an upper-tailed hypothesis test an upper or lower confidence bound will be constructed, respectively, rather than a confidence interval.
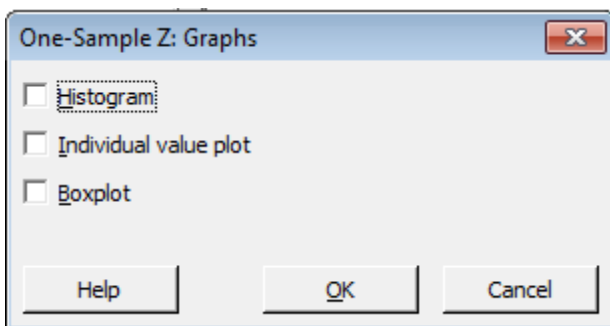
In the **One-Sample Z: Graphs** dialog box choose to display a histogram, individual data plot, and/or a boxplot of the variables. The graphs show the sample mean and a confidence interval (or bound) for the mean. When you do a hypothesis test , the graphs also show the null hypothesis test value.

Histogram of Yards

Histogram of Yards
(with Ho and 95% Z-confidence interval for the Mean, and StDev = 12)
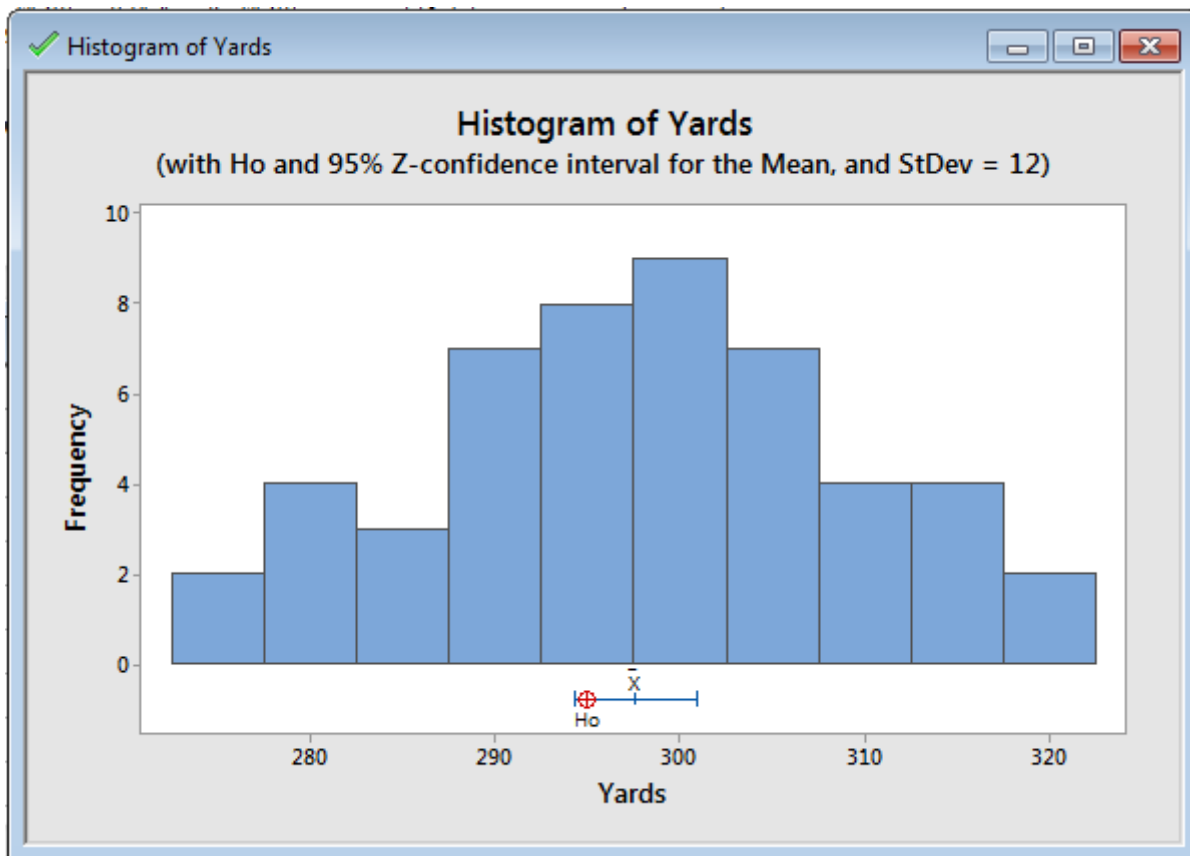
Session

One-Sample Z: Yards

```
Test of μ = 295 vs ≠ 295
The assumed standard deviation = 12


Variable   N    Mean   StDev  SE Mean       95% CI          Z     P
Yards      50  297.60  11.30    1.70  (294.27, 300.93)  1.53  0.126
```

When you have summary values for the sample size and mean, choose the **Summarized data:** option. You will be prompted to enter the value for the Sample size: and Sample mean:
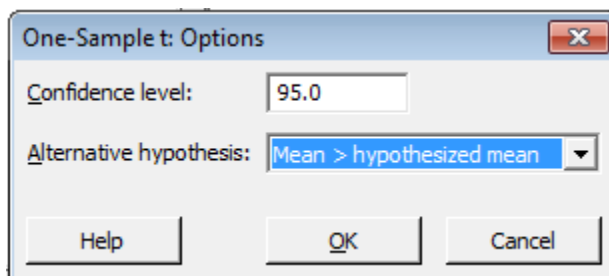
ONE-SAMPLE T FOR THE MEAN

The Stat > Basic Statistics > 1-Sample t for the Mean computes a confidence interval or performs a hypothesis test of the mean when σ is unknown. This procedure is based upon the t-distribution, which is derived from a normal distribution with unknown σ. For small samples, this procedure works best if your data were drawn from a distribution that is normal or close to normal. This procedure is more conservative than the Z-procedure and should always be chosen over the Z-procedure with small sample sizes and an unknown σ. Many analysts choose the t-procedure over the Z-procedure anytime σ is unknown.

For a two-tailed one-sample t, $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ where μ is the population mean and μ_0 is the hypothesized population mean.
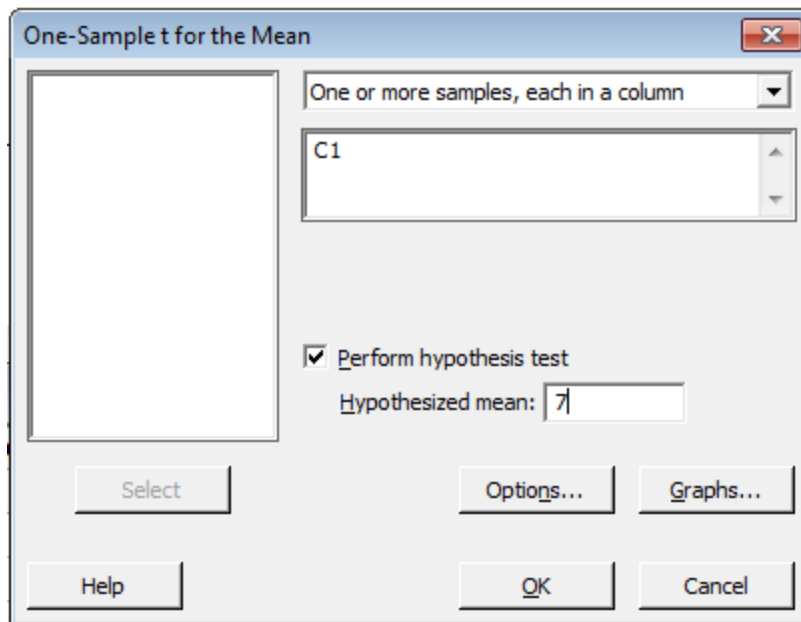
Choose **One or more samples, each in a column:** if you have entered raw data in columns. Enter the columns containing the sample data. Enter the population known (σ) standard deviation. Check **Perform hypothesis test:** to perform the hypothesis test**.** Enter the **Hypothesized mean:** μ_0. Choose the Options… command.

In the **One-Sample t: Options** dialog box specify the confidence level for the confidence interval, or define the alternative hypothesis. Choose *Mean < hypothesized mean* (lower-tailed), *Mean ≠ hypothesized mean* (two-tailed), or *Mean > hypothesized mean* (upper-tailed). If you choose a lower-tailed or an upper-tailed hypothesis test an upper or lower confidence bound will be constructed, respectively, rather than a confidence interval.
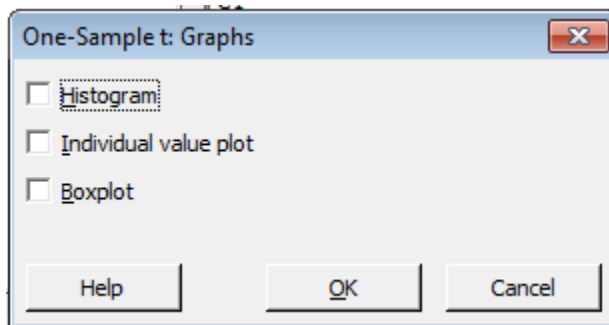
In the **One-Sample t: Graphs** dialog box choose to display a histogram, individual data plot, and/or a boxplot of the variables. The graphs show the sample mean and a confidence interval (or bound) for the mean. When you do a hypothesis test , the graphs also show the null hypothesis test value.

Histogram of Rating

Histogram of Rating
(with Ho and 95% t-confidence interval for the mean)
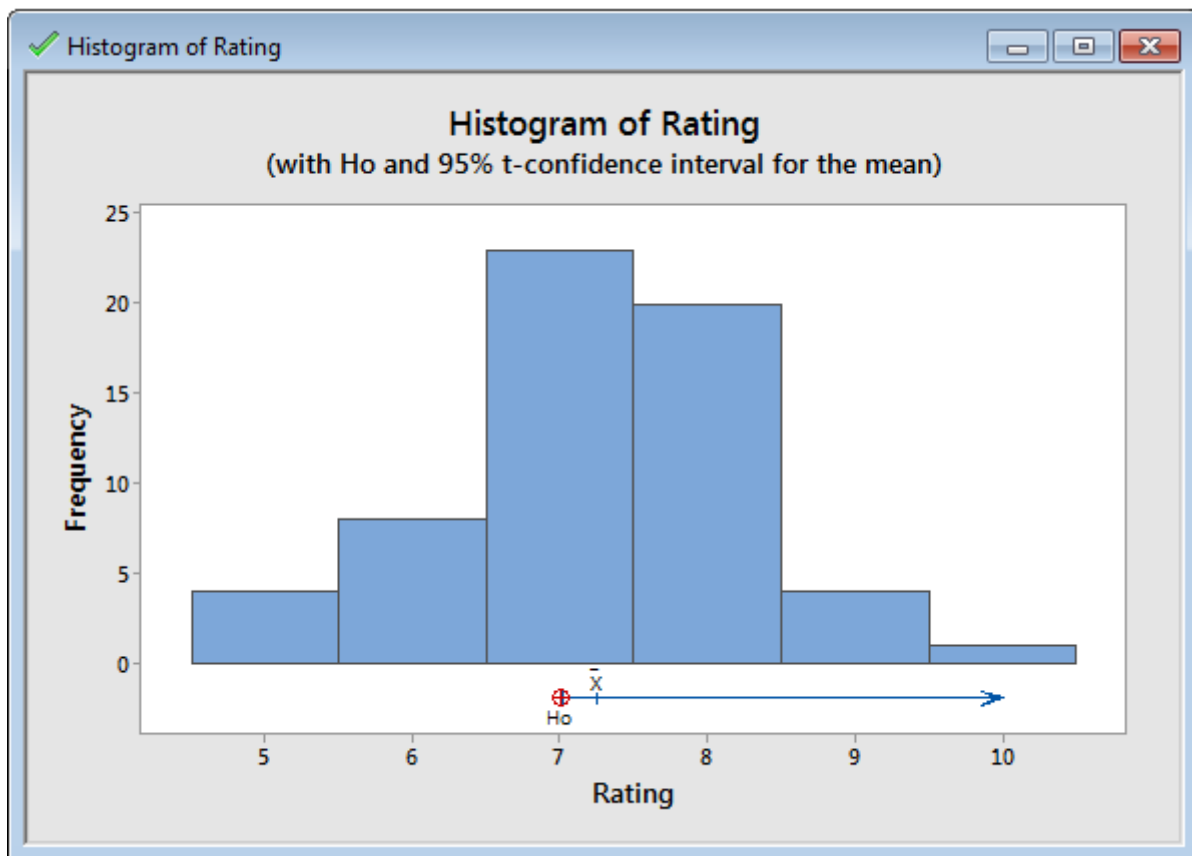
Session

One-Sample T: Rating

Test of $\mu = 7$ vs > 7

Variable	N	Mean	StDev	SE Mean	95% Lower Bound	T	P
Rating	60	7.250	1.052	0.136	7.023	1.84	0.035

TWO-SAMPLE T FOR THE MEAN

The Stat > Basic Statistics > 2-Sample t for the Mean option computes a confidence interval and performs a hypothesis test of the difference between two population means when s's are unknown and samples are drawn independently from each other. This procedure is based upon the t-distribution, and for small samples it works best if data were drawn from distributions that are normal or close to normal. You can have increasing confidence in the results as the sample sizes increase. When you have dependent samples, use Stat > Basic Statistics > Paired t.

For a two-tailed 2-sample t, $H_0: \mu_1 - \mu_2 = \delta_0$ versus $H_1: \mu_1 - \mu_2 \neq \delta_0$ where μ_1 and μ_2 are the population means and δ_0 is the hypothesized difference between the two population means.

DIALOG BOX ITEMS
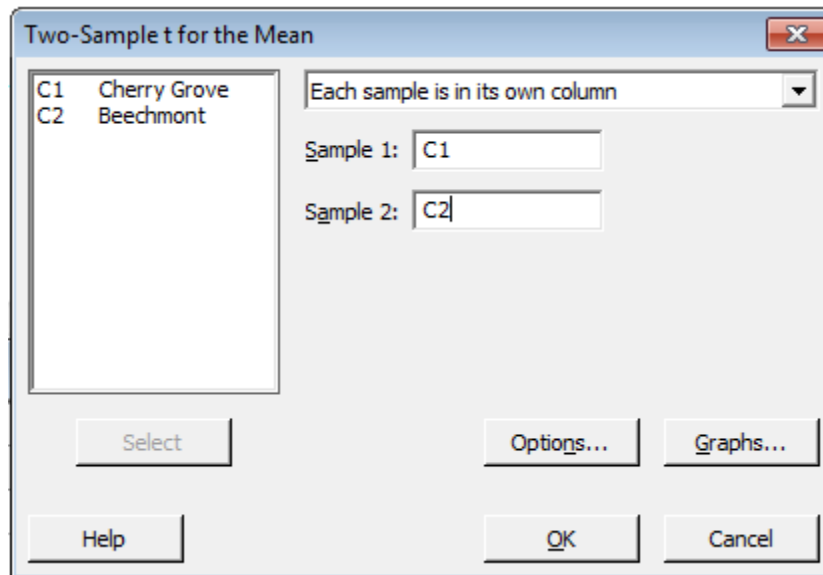
In the dropdown list box, choose the data orientation as: *Both samples are in one column, Each sample is in its own column,* or *Summarized data*.

Choose **Both samples are in one column:** if the sample data are in a single column, differentiated by sample IDs in a second column. Enter the column containing the data in the Samples: text box. Enter the column containing the sample IDs in the Sample IDs: text box.

Choose **Each sample is in its own column:** if the data of the two samples are in separate columns. Enter the column containing one sample in the Sample 1: text box. Enter the column containing the other sample in the Sample 2: text box.

Choose **Summarized data:** if you have summary values for the sample size, mean, and standard deviation for each sample. Enter the Sample 1 data in the Sample size, Sample mean, and Standard deviation text boxes. Enter the Sample 2 data in the Sample size, Sample mean, and Standard deviation text boxes.
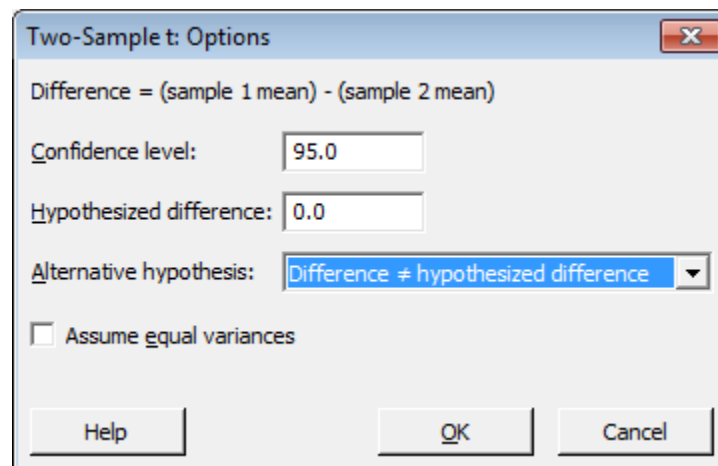
In the **Two-Sample t: Options** dialog box, enter the level of confidence desired as a number between 0 and 100. The default is 95. Enter the null hypothesis value in the Hypothesized difference: text box, which is the hypothesized difference in population means, δ_0 . The default is zero, or that the two population means are equal. In the Alternative hypothesis: dropdown list box, choose Difference < hypothesized difference (lower-tailed), Difference   hypothesized difference (two-tailed), or Difference > hypothesized difference (upper-tailed) depending on the kind of test that you want. Check the Assume equal variances checkbox to assume that the populations have equal variances. The default is to assume unequal variances.
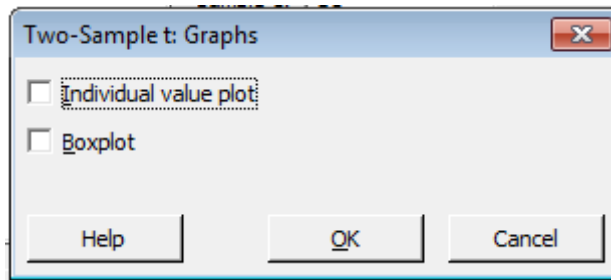
In the **Two-Sample t: Graphs** dialog box, choose the Individual value plot check box to display an individual value plot for each sample. Choose the Boxplot check box to display a boxplot for each sample.

PAIRED T FOR THE MEAN

The Stat > Basic Statistics > Paired t for the Mean options performs a paired t-test. When data are paired, as with before-and-after measurements, the paired t-procedure results in a smaller variance and greater power of detecting differences than would the 2-sample t-procedure, which assumes that the samples were independently drawn. This is appropriate for testing the mean difference between paired observations when the paired differences follow a normal distribution. The Paired t command is used to compute a confidence interval and perform a hypothesis test of the mean difference between paired observations in the population. A paired t-test matches responses that are dependent or related in a pairwise manner. This matching allows you to account for variability between the pairs usually resulting in a smaller error term, thus increasing the sensitivity of the hypothesis test or confidence interval.

Typical examples of paired data include measurements on twins or before-and-after measurements. For a paired t-test: H0: m d = m 0   versus   H1: m d ≠ m 0 where m d is the population mean of the differences and m 0 is the hypothesized mean of the differences. When the samples are drawn independently from two populations, use Stat > Basic Statistics > 2-sample t.
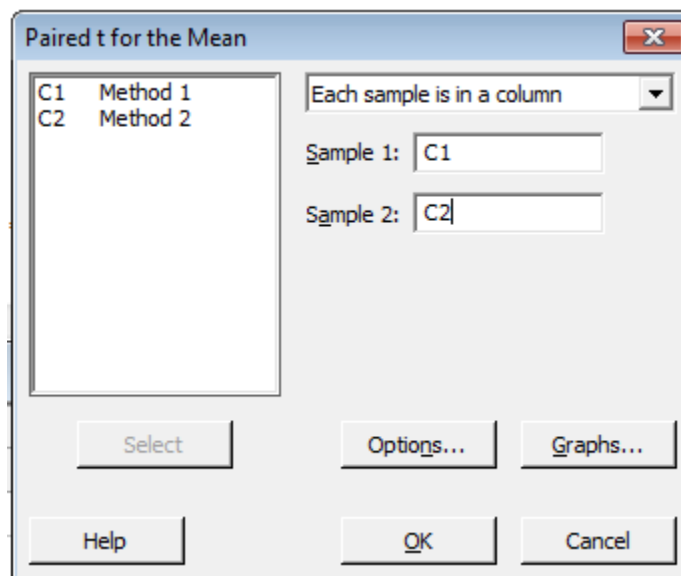

DIALOG BOX ITEMS

In the dropdown list box, choose the data orientation as: *Each sample is in its own column,* or *Summarized data*.

Choose Each sample is in a column: if you have entered raw data in two columns. In the Sample 1: text box enter the column containing the first sample. Enter the column containing the second sample in the Sample 2: text box.

Choose Summarized data (differences): if you have summary values for the sample size, mean, and standard deviation of the difference. Enter the value for the sample size, sample mean, and sample standard deviation in the appropriate text boxes..

In the **Paired t: Options** dialog box, enter the level of confidence desired as a number between 0 and 100. The default is 95. Enter the null hypothesis value in the Hypothesized difference: text box, which is the hypothesized difference in population means, δ_0 . The default is zero, or that the two population means are equal. In the Alternative hypothesis: dropdown list box, choose Difference < hypothesized difference (lower-tailed), Difference ≠ hypothesized difference (two-tailed), or Difference > hypothesized difference (upper-tailed) depending on the kind of test that you want.
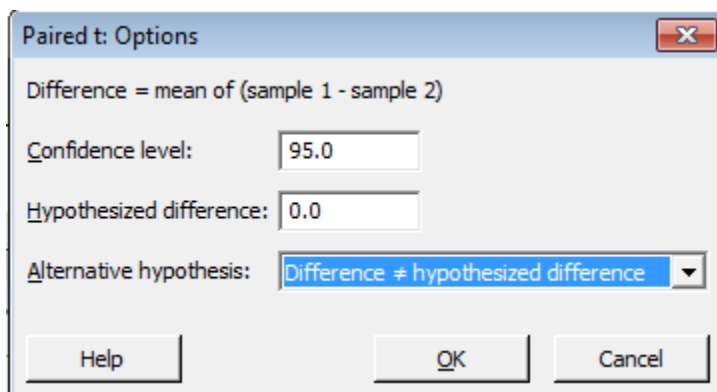
In the **Paired t: Graphs** dialog box choose to display a histogram, individual data plot, and/or a boxplot of the variables. The graphs show the sample mean and a confidence interval (or bound) for the mean. When you do a hypothesis test , the graphs also show the null hypothesis test value.
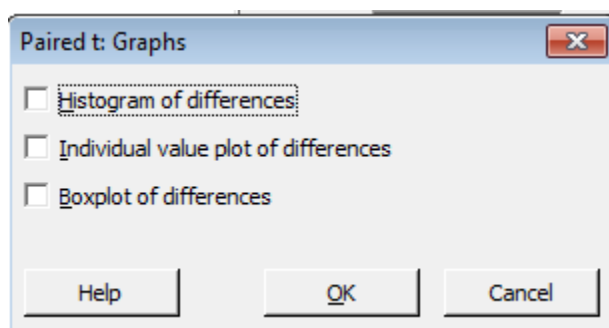
1 PROPORTION

The Stat > Basic Statistics > 1 Proportion option is used to compute a confidence interval and perform a hypothesis test of the proportion . For example, Pine Creek golf course is interested in increasing the proportion of women players and has implemented a special promotion designed to attract women golfers. One month after the promotion was implemented; the course manager would like to determine if the proportion of women players has increased.

For a two-tailed 1 proportion test, H_0: $p = p_0$ versus H_1: $p \neq p_0$ where p is the population proportion and p_0 is the hypothesized value.

In the dropdown list box, choose the data orientation as *One or more samples, each in a column* or *Summarized data*.

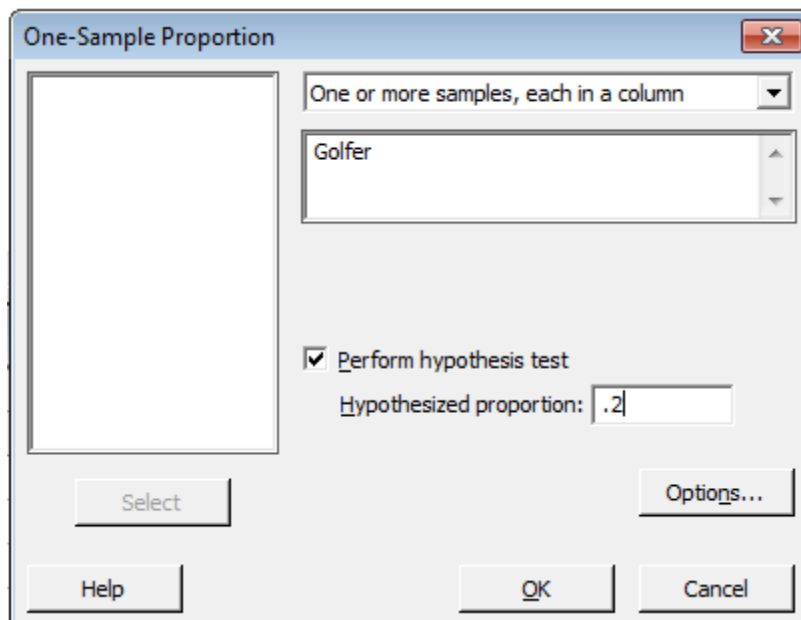Choose *One or more samples, each in a column* if you have data in columns, then enter the columns containing the sample data. Each cell of these columns must be one of two possible values and correspond to one item or subject. The possible values in the columns must be identical if you enter multiple columns.

Choose *Summarized data* if you have summary values for the numbers of trials and events. Enter the Number of events in the textbox as the number of observed events. If you enter more than one value; the single integer you enter in Number of trials will apply to them all. Enter the Number of trials in the textbox as a single value for the number of trials.

Check Perform hypothesis test to perform the hypothesis test that the population proportion is equal to a specified value. Enter the Hypothesized proportion as the value of the proportion for the test's null hypothesis.

In the **One-Sample Proportion: Options** dialog box, enter the level of confidence desired as a number between 0 and 100. The default is 95. From the Alternative hypothesis: dropdown list, choose the form of the alternative hypothesis: Proportion < hypothesized value (lower-tailed), Proportion ≠ hypothesized value (two-tailed), or Proportion > hypothesized value (upper-tailed). Note that if you choose a lower-tailed or an upper-tailed hypothesis test, an upper or lower confidence bound will be constructed, respectively, rather than a confidence interval. From the Method: dropdown list, choose either the exact method or the normal approximation to the binomial distribution method for calculating the hypothesis test and confidence interval. Most textbooks use the normal approximation method because it is easy for students to calculate by hand.
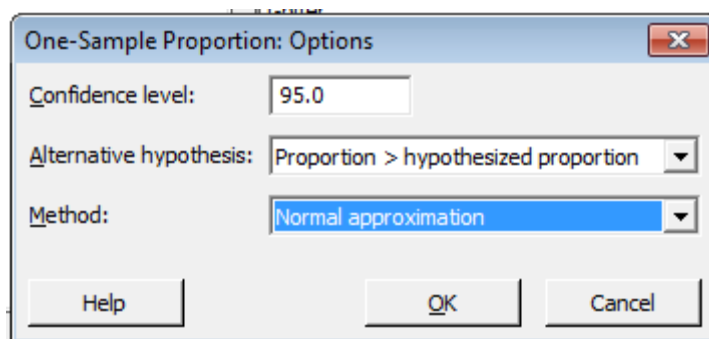
Test and CI for One Proportion: Golfer

Test of p = 0.2 vs p > 0.2

Event = Male

| Variable	X	N	Sample p	95% Lower Bound	Z-Value	P-Value
Golfer	300	400	0.750000	0.714388	27.50	0.000

Using the normal approximation.

NOTE: The results of this output indicate that the Event in question is Male.

---

2 PROPORTIONS

The Stat > Basic Statistics > 2 Proportions option is used to compute a confidence interval and perform a hypothesis test of the difference between two proportions. Minitab offers two hypothesis tests for the difference between two proportions: Fisher's exact test , and a test based on a normal approximation. The normal approximation test may be inaccurate for samples in which the number of events in either sample is less than five, or if the difference between the numbers of trials and events in either sample is less than five. Fisher's exact test is accurate for all sample sizes , but can only be calculated when the null hypothesis states that the population proportions are equal. In other words, Minitab only performs Fisher's exact test when you specify a test difference of zero in the Options subdialog box.

For example, suppose a tax preparation firm is interested in comparing the quality of work at two of its regional offices. The firm proceeds to estimate the proportion of erroneous returns prepared at each office. For a two-tailed test of two proportions: H_0: p_1 - p_2 = p_0 versus H_1: p_1 - p_2 ≠ p_0 where p_1 and p_2 are the proportions of events in populations 1 and 2, respectively, and p0 is the hypothesized difference between the two proportions.
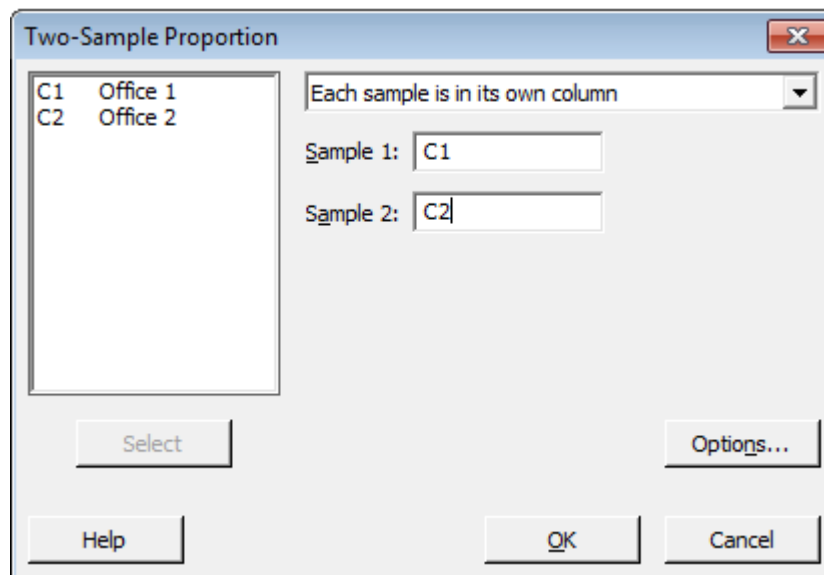
DIALOG BOX ITEMS

In the dropdown list box, choose the data orientation as *Both samples are in one column*, *Each sample is in its own column*, or *Summarized data*.

Choose *Both samples are in one* column if you have entered raw data into a single column with a second column of subscripts identifying the sample. In the Samples: textbox enter the column containing the raw data (in the following example, Call time). In the Sample IDs: textbox, enter the column containing the sample subscripts (Operator).

Used to identify or categorize values in another column. In the example below, the subscript column, Operator, identifies which values in the response column, Call time, are associated with each operator.

Call time	Operator
12.5	A
11.2	A
11.9	A
12.2	A
8.2	B
8.4	B
7.6	B
8.5	B

Choose *Each sample is in its own column* if you have entered raw data into single columns for each sample. In the Sample 1: textbox, enter the column containing the raw data for the first sample. In the Sample 2: textbox, enter the column containing the raw data for the second sample.

Choose *Summarized data* if you have summary values for the numbers of trials and events. In the textboxes provided for Sample 1, enter the number of events in the first sample and the number of trials in the first sample. In the textboxes provided for Sample 2, enter the number of events in the first sample and the number of trials in the first sample.

In the Two-Sample Proportion: Options dialog box, enter the level of confidence desired as a number between 0 and 100. The default is 95. Enter the hypothesized difference (typically 0). From the Alternative hypothesis: dropdown list, choose the form of the alternative hypothesis: Difference < hypothesized difference (lower-tailed), Difference ≠ hypothesized difference (two-tailed), or Difference > hypothesized difference (upper-tailed).Note that if you choose a lower-tailed or an upper-tailed hypothesis test, an upper or lower confidence bound will be constructed, respectively, rather than a confidence interval. From the Test method: dropdown list, choose to estimate the proportions separately or to use a pooled estimate of p for the hypothesis test.

Two-Sample Proportion: Options	⊠
Difference = (sample 1 proportion) - (sample 2 proportion)	
Confidence level:	95.0
Hypothesized difference:	0.0
Alternative hypothesis:	Difference ≠ hypothesized difference ▼
Test method:	Estimate the proportions separately ▼
Help	OK Cancel

## Test and CI for Two Proportions: Office 1, Office 2

```
Event = Yes


Variable    X    N  Sample p
Office 1   35  250  0.140000
Office 2   21  250  0.084000


Difference = p (Office 1) - p (Office 2)
Estimate for difference:  0.056
95% CI for difference:  (0.000933173, 0.111067)
Test for difference = 0 (vs ≠ 0):  Z = 1.99  P-Value = 0.046

Fisher's exact test: P-Value = 0.064
```

1 VARIANCE

The Stat > Basic Statistics > 1 Variance option is used to calculate confidence intervals for the standard deviation and variance of a population, and perform a hypothesis test to determine whether the population variance equals a specified value. If you enter summary values, Minitab analyzes one sample at a time and cannot calculate the Bonett method for use with nonnormal data.
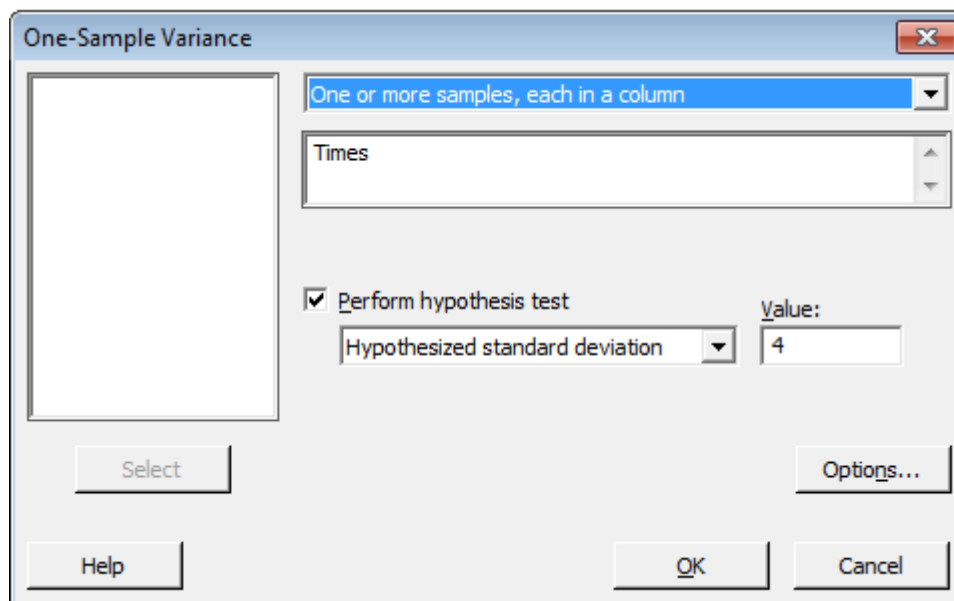
DIALOG BOX ITEMS

In the dropdown list box, choose the data orientation as *One or more samples, each in a column*, *Sample standard deviation*, or *Sample variance.*

Choose *One or more samples, each in a column* if you entered raw data into columns and enter the column(s) that contain the data. If you enter multiple columns, Minitab performs separate 1-sample analyses on each column.

Choose *Sample standard deviation* if you have summary values for a sample's size and standard deviation. Enter the sample size in the Sample size: text box. Enter the sample's standard deviation in the Sample standard deviation: textbox.

Choose *Sample variance* if you have summary values for a sample's size and variance. Enter the sample size in the Sample size: text box. Enter the sample's variance in the Sample variance: text box.

Check the Perform hypothesis test: checkbox to perform the hypothesis test that the population standard deviation or variance equals a specified value.

In the associated dropdown list box, choose *Hypothesized standard deviation* to test the standard deviation. Choose *Hypothesized variance* to test the variance. In the Value: textbox, enter the hypothesized value for standard deviation or variance.

In the One-Sample Variance: Options dialog box, specify the confidence level as a number between 0 and 100, exclusive. The default is 95. From the Alternative hypothesis: dropdown list, choose Standard deviation < hypothesized standard deviation (lower-tailed), Standard deviation ≠ hypothesized standard deviation (two-tailed), or Standard deviation > hypothesized standard deviation (upper-tailed). If you choose a lower-tailed or an upper-tailed hypothesis test, an upper or lower confidence bound will be constructed, respectively, rather than a confidence interval.
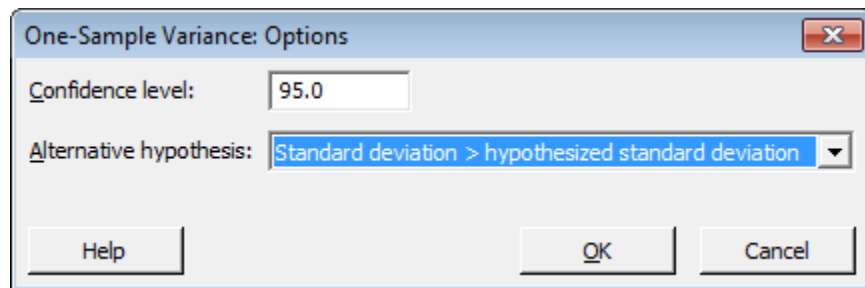
Test and CI for One Variance: Times

Method

```
Null hypothesis        σ = 4
Alternative hypothesis  σ > 4
```

The chi-square method is only for the normal distribution.
The Bonett method is for any continuous distribution.

Statistics

Variable	N	StDev	Variance
Times	24	2.21	4.90

95% One-Sided Confidence Intervals

Variable	Method	Lower Bound for StDev	Lower Bound for Variance
Times	Chi-Square	1.79	3.20
	Bonett	1.76	3.08

Tests

Variable	Method	Test Statistic	DF	P-Value
Times	Chi-Square	7.04	23	0.999
	Bonett	—	—	1.000

2 VARIANCES

The Stat > Basic Statistics > 2 Variances option is used to perform hypothesis tests and compute confidence intervals for the ratios between two populations' variances and standard deviations ; a ratio of 1 suggests equality between populations. Use this test to determine if one treatment condition has more variability than the other. For example, a school district may be considering renewing its school bus service for the coming year and must select one of two bus companies. If the variances of arrival times associated with the two services are equal, the school district will select the company with the better financial terms.

By default, Minitab uses two tests: Bonett's test and Levene's test. For each test, the null hypothesis states that the two variances are equal ($H_0: \sigma_1^2/ \sigma_2^2 = 1$). The alternative hypothesis can be left-tailed ($H_1: \sigma_1^2/ \sigma_2^2 < 1$), right-tailed ($H_1: \sigma_1^2/ \sigma_2^2 > 1$), or two-tailed ($H_1: \sigma_1^2/ \sigma_2^2 \neq 1$). Optionally, test ratios other than 1 (equality) can be specified.
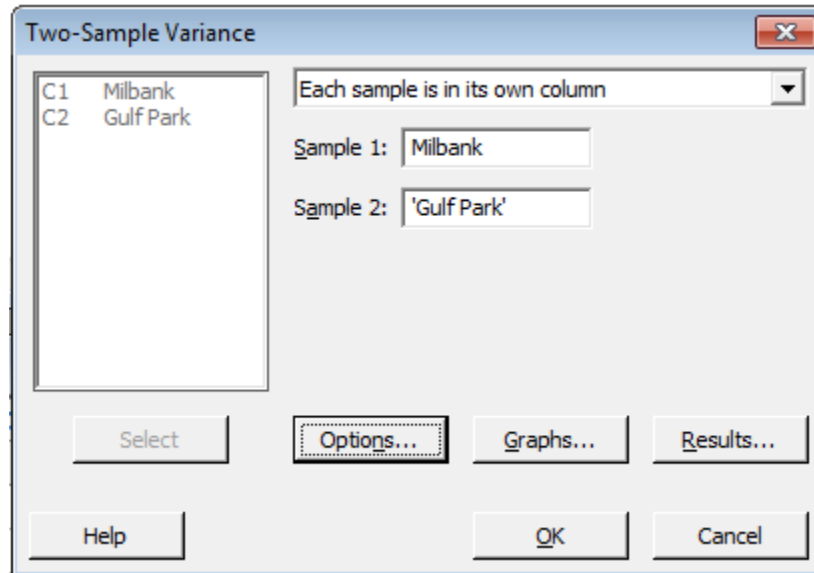
DIALOG BOX ITEMS

In the dropdown list box, choose the data orientation as *Both samples are in one column*, *Each sample is in its own column*, *Sample standard deviations* or *Sample variances.*

Choose *Both samples are in one column* if the sample data are in a single column, and are differentiated by sample IDs in a second column. Enter the column that contains the data for both samples in the Samples: textbox. Enter the column that indicates which sample each observation belongs to in the Sample IDs: textbox.

Choose *Each sample is in its own column* if the data for each sample is in a different column. Enter the column that contains the data for the first sample in the Sample 1: textbox. Enter the column that contains the data for the second sample in the Sample 2: textbox.

Choose *Sample standard deviations* to enter summary values for the sample size and standard deviation of each sample. For summary data, only the F-test can be performed. The F-test is accurate only for normally distributed data. Enter the sample size for each sample in the Sample size: textboxes. Enter the standard deviation for each sample in the Sample variance: textboxes.

Choose *Sample variances* to enter summary values for the sample size and variance of each sample. For summary data, only the F-test can be performed. The F-test is accurate only for normally distributed data. Enter the sample size for each sample in the Sample size: textboses. Enter the variance for each sample in the Variance: textbox.

In the Two-Sample Variance: Options dialog box, specify the type of ratio, the confidence level, the hypothesized ratio, and the alternative hypothesis for the test. You can also choose to use the F-test rather than Levene's and Bonett's test. In the Ratio dropdown list box, choose to test the ratio of the standard deviations or the ratio of the variances. Enter the confidence level in the Confidence level: textbox as any number between 0 and 100. The default is 95. The significance level for the tests (also called alpha or α) is also determined by the confidence level. The significance level = 1 − (confidence level / 100). For example, if you choose 95 as the confidence level, the significance level for the tests is 1 − (95 / 100) = 0.05. Enter the hypothesized value for the ratio in the Hypothesized ratio: textbox. Use the default value of 1 to test the equality of two population variances or standard deviations. In the Alternative hypothesis: dropdown list box, choose the alternative hypothesis that you want to test. To use the F-test rather than Levene's and Bonett's test, check the use test and confidence intervals based on normal distribution checkbox. The F-test is accurate only for normally distributed data. Any departure from normality can cause the F-test to yield inaccurate results.

Two-Sample Variance: Options

Ratio: (sample 1 variance) / (sample 2 variance)

Confidence level: 95.0

Hypothesized ratio: 1

Alternative hypothesis: Ratio ≠ hypothesized ratio

☑ Use test and confidence intervals based on normal distribution

Help   OK   Cancel
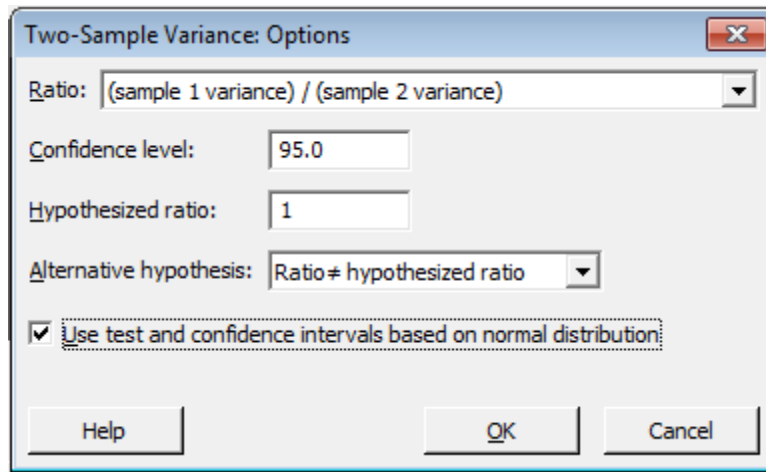
In the Two-Sample Variance: Graphs dialog box, display graphs to help you visualize the test results and the shape and spread of the two samples' distributions. Check the Summary plot: checkbox to display a summary of the results. The summary includes confidence intervals for the ratio of the standard deviations or variances and p-values for the tests. The summary also includes a confidence interval for the standard deviation or variance of each sample and boxplots of each sample. Check the Histogram: checkbox to display a histogram for each sample. Check the Individual value plot: checkbox to display an individual value plot for each sample.

Two-Sample Variance: Graphs

☑ Summary plot
☐ Histogram
☐ Individual value plot

Help   OK   Cancel
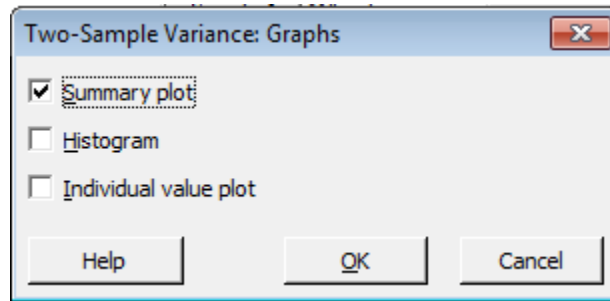
In the Two-Sample Variance: Results dialog box, choose the results to display. Check the Method: checkbox to display the method information, including the null hypothesis, the alternative hypothesis, and the significance level (also called a). Check the Statistics: checkbox to display descriptive statistics including the standard deviation of each sample, the variance of each sample, and the confidence interval for the standard deviation or the variance of each sample. The ratio of the standard deviations and the ratio of the variances are also displayed. Check the Confidence intervals: checkbox to display confidence intervals for the ratio of the standard deviations and the ratio of the variances. Check the Test: checkbox to display the results of the test, including the degrees of freedom, the test statistics, and the p-values. When the sample sizes are unequal, the test statistic for Bonett's method is undefined. However the p-value can be calculated by inverting the confidence interval procedure.

Two-Sample Variance: Results

☑ Method
☑ Statistics
☑ Confidence intervals
☑ Test

Help   OK   Cancel
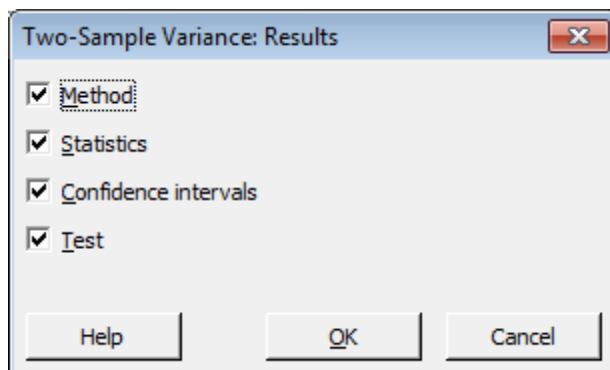
Test and CI for Two Variances: Milbank, Gulf Park

Method

Null hypothesis   Variance(Milbank) / Variance(Gulf Park) = 1
Alternative hypothesis Variance(Milbank) / Variance(Gulf Park) ≠ 1
Significance level   α = 0.05

F method was used. This method is accurate for normal data only.

Statistics

| | | | | 95% CI for |
Variable	N	StDev	Variance	Variances
Milbank	26	6.930	48.021	(29.536, 91.505)
Gulf Park	16	4.472	20.000	(10.914, 47.907)

Ratio of standard deviations = 1.550
Ratio of variances = 2.401

95% Confidence Intervals

| | CI for StDev | CI for Variance |
Method	Ratio	Ratio
F	(0.945, 2.406)	(0.893, 5.789)

Tests

| | | | Test | |
Method	DF1	DF2	Statistic	P-Value
F	25	15	2.40	0.081

LINEAR REGRESSION

Regression analysis is used to investigate and model the relationship between a response variable and one or more predictors. Minitab provides the following regression procedures:

- Least squares : Use when your response variable is continuous .

- Nonlinear regression: Use when you cannot adequately model the relationship with linear parameters.

- Orthogonal regression: Use when the response and predictor both contain measurement error.

- Partial least squares regression : Use when your predictors are highly correlated or outnumber your observations.

- Logistic regression: Use when your response variable is categorical.

- Poisson regression: Use when your response variable counts occurrences.

Both least squares and logistic regression methods estimate parameters in the model so that the fit of the model is optimized. Least squares methods minimize the sum of squared errors to obtain parameter estimates, whereas Minitab's logistic regression obtains maximum likelihood estimates of the parameters. See Generalized Linear Models Overview for more information. Partial least squares (PLS) extracts linear combinations of the predictors to minimize prediction error. See Partial Least Squares Overview for more information.

Use the table below to select a procedure:

Use...	To...	Response type	Estimation method
Regression	perform simple, multiple, polynomial, and stepwise least squares regression with continuous and categorical predictors	continuous	least squares
Best Subsets	identify subsets of the predictors based on the maximum R criterion	continuous	least squares
Fitted Line Plot	perform linear and polynomial regression with a single predictor and plot a regression line through the data	continuous	least squares
Nonlinear Regression	perform simple or multiple regression using the nonlinear function of your choice	continuous	least squares
Stability Study	analyze the stability of a product over time and determine the product's shelf life	continuous	least squares, Restricted Maximum Likelihood (REML)
Orthogonal Regression	perform orthogonal regression with one response and one predictor	continuous	orthogonal
PLS	perform regression with ill-conditioned data	continuous	biased, non-least squares
Binary Logistic	perform logistic regression on a response with only two possible values, such as presence or absence	categorical	maximum likelihood
Ordinal Logistic	perform logistic regression on a response with three or more possible values that have a natural order, such as none, mild, or severe	categorical	maximum likelihood
Nominal Logistic	perform logistic regression on a response with three or more possible values that have no natural order, such as sweet, salty, or sour	categorical	maximum likelihood
Poisson Regression	perform Poisson regression on a response that counts occurrences	discrete	maximum likelihood

Stat > Regression > Regression > Fit Regression Model

Use Regression to fit least squares models when you have continuous and/or categorical predictors. You can:

- fit interaction and polynomial terms

- perform stepwise regression

- store regression statistics

- examine residual diagnostics

- perform the pure error lack-of-fit test when your data contain replicates

- transform highly skewed data

The default model contains the variables that you enter in **Continuous predictors** and **Categorical predictors**. If you want to add interaction and polynomial terms, use the tools in the **Model** subdialog box.

Minitab stores the last model that you fit for each response variable. You can use the stored models to quickly generate predictions, contour plots, surface plots, overlaid contour plots, factorial plots, and optimized responses.

See Stored Model Overview for a discussion about how to use stored models.

DIALOG BOX ITEMS

Responses: Select the continuous variable(s) that you want to explain or predict with the predictors (X). The response is also called the Y variable. If there is more than one response variable, Minitab fits separate models for each response.

Continuous predictors: Select the continuous variables that explain changes in the response. The predictor is also called the X variable.

Categorical predictors: Select the categorical classifications or group assignments, such as a type of raw material, that explain changes in the response. The predictor is also called the X variable.

<Model>

Stat > Regression > Regression > Fit Regression Model > Model

Use to add interaction terms and polynomial terms to your model. By default, the model contains only the predictor variables that you entered in the main dialog box. Click **Default** to return to this model at any time.

To add terms to the model, select at least one predictor or term. To select multiple items, press the **Ctrl** key while you click the predictors and/or terms.

There are three basic ways you can add terms. Under **Dialog box items**, we use examples to illustrate the ways. For the examples, assume the predictor list has 3 continuous variables, X, Y, Z and 2 categorical variables, A, B.

See Tips for Creating Terms for more information.

Add terms using selected predictors and model terms:

Interactions through order: This option adds all interactions through the specified order. Suppose you select predictors X, Y, A and interactions through order 3. When you click **Add**, Minitab adds X*Y, X*A, Y*A, X*Y*A.

Terms through order: Use to model curvature . This option adds powers and interactions through the specified order. Powers are for continuous predictors. Suppose you select X, Y, A and terms through order 3. When you click **Add**, Minitab adds the power terms for X and Y: X*X, Y*Y, X*X*X, Y*Y*Y. Minitab also adds interactions for the predictor variables and powers: X*Y, X*A, Y*A, X*X*Y, X*Y*Y, X*X*A, X*Y*A, Y*Y*A.

Cross predictors and terms in the model: This option can be used in several ways.

1   You can cross two or more predictors. Suppose you select X, Y, Z. When you click **Add**, Minitab adds X*Y*Z.

2   You can cross two or more terms that are already in the model. Suppose X*A and X*B are in the model. If you select only these terms and click **Add**, Minitab adds X*X*A*B.

3   You can cross predictors with terms in the model. Suppose X*X and Y*Y are in the model. If you select these terms and predictors A, B then click **Add**, Minitab adds X*X*A, X*X*B, Y*Y*A, Y*Y*B,   Each predictor is crossed with each model term, but the predictors are not crossed with themselves and the model terms are not crossed with themselves.

Default: Populates the model with only the predictor variables that you entered in the main dialog box.

Delete Terms: You can delete one or more terms from the model. Select the terms and click the **Delete** button in the dialog. You can also double-click a term to delete it.

Reorder Terms: When you add terms, they are put at the end of the list of model terms. To move a term, select it then click one of the arrow buttons in the dialog to move the term up or down. You can also move a contiguous block of terms. You can also move a contiguous block of terms. Click the first term then hold the **Shift** key and click the last term to select the whole block. Then click the appropriate arrow to move the block.

Include the constant term in the model: Check to fit the model with a y-intercept (constant term) for the regression line . Consider fitting a model without the constant only if you remove all insignificant terms from the model and the constant is not significantly different from 0. If you do not include the constant in the correct circumstance, this can improve the fit of the model and the precision of the predicted values.

<Options>

Stat > Regression > Regression > Fit Regression Model > Options

Use to:

- Perform a weighted regression

- Specify the confidence level and type of confidence interval

- Choose the type of sum of squares

- Perform a Box-Cox transformation on your response data when the residuals are not normally distributed or do not have constant variance

When you transform your data, Minitab transforms the response data and uses it in the analysis. Under most conditions, it is not necessary to correct for nonnormality unless the data are highly skewed. When you use Box-Cox, all response data must be greater than 0. Check your model carefully before using the Box-Cox transformation.

Note   Minitab cannot calculate the optimal λ (lambda) when you use a stepwise procedure.

DIALOG BOX ITEMS

Weights: Enter a numeric column of weights to perform weighted regression. Weights must be greater than or equal to zero. The length of the weights column must match the length of the response column.

Confidence level for all intervals: Enter the confidence level. The default is 95.

Type of confidence interval: Choose the type of confidence interval: two-sided (default), lower bound, or upper bound.

Sums of Squares for Tests: Choose a sums of squares for calculating F and p-values. Typically, use the adjusted SS. Use sequential SS to determine the significance of terms by the order that they enter the model.

Adjusted (Type III): Measures the reduction in the SS for each term relative to a model that contains all of the remaining terms.

Sequential (Type I): Measures the reduction in the SS when a term is added to a model that contains only the terms before it.

Box-Cox Transformation

No transformation: Choose to use your original response data.

Optimal λ (lambda): Choose to have Minitab search for an optimal value.

λ = 0 (natural log): Choose to use the natural log of the data.

λ = 0.5 (square root): Choose to use the square root of the data.

λ **:** Choose to transform the data using another λ value. Enter a value between -5 and 5.

<Coding>

Stat > Regression > Regression > Fit Regression Model > Coding

Use to:

- Specify the coding scheme for categorical predictors

- Specify the reference level for each categorical predictor

- Standardize the continuous predictors

Standardizing the continuous predictors can improve the interpretation of the model for specific conditions. You can standardize the continuous predictors using the following methods:

- Center the continuous predictors by subtracting the mean: This method helps reduce multicollinearity , which improves the precision of the coefficient estimates. This method is helpful when your model contains highly correlated predictors, higher-order terms, and interaction terms. Each coefficient represents the expected change in the response given a one unit change in the predictor, using the original measurement scale.

- Adjust the scale of the continuous predictors by dividing by the standard deviation: This method allows you to compare the size of the coefficients because they use a comparable scale. This approach is helpful when you want to know which predictors have a larger effect, while controlling for differences in scale. However, each coefficient represents the expected change in the response given a change of one standard deviation in the predictor.

DIALOG BOX ITEMS

Coding for categorical predictors: Use when you have categorical predictors. To perform the analysis, Minitab needs to recode the categorical data. Base your decision on whether you want to compare the levels to the overall mean or the mean of a reference level.

(− 1, 0, +1): Choose to estimate the difference between each level mean and the overall mean.

(1, 0): Choose to estimate the difference between each level mean and the reference level's mean.

Categorical predictor reference level: If you choose the (1, 0) coding scheme, the reference level table becomes active in the dialog box. Minitab compares the means of the nonreference level(s) to the reference level. Changing the reference level does not affect the overall significance, but it can make the results more meaningful.

Categorical predictor: Shows all the names of categorical predictors in your model. This column does not take any input.

Reference level: Choose a reference level for each categorical predictor.

Standardize continuous predictors: Use to control whether Minitab standardizes the continuous predictors. The standardized predictors are only used to fit the model and are not stored in the worksheet.

Do not standardize: Choose to use your original data for the continuous predictors.

Subtract the mean, then divide by the standard deviation: Choose to both center the predictors and to place them on a comparable scale.

Subtract the mean: Choose to center the predictors.

Divide by the standard deviation: Choose to use a comparable scale for all predictors.

Subtract a specified value, then divide by another: Choose to specify values rather than the mean and standard deviation estimates from the sample.

Continuous predictor: Shows all of the names of continuous predictors in your model. This column does not take any input.

Subtract: Type the value to subtract from each continuous predictor.

Divide by: Type the value that Minitab uses to divide the result of the subtraction.

Specify low and high levels to code as -1 and +1: Choose to transform the data linearly. All data values that fall between the **Low** and **High** values that you specify are coded to fall between -1 and +1. Designed experiments (DOE) use this scheme.

Continuous predictor: Shows all of the names of continuous predictors in your model. This column does not take any input.

Low: Type a value to code as -1. The default is the minimum value in the sample.

High: Type a value to code as +1. The default is the maximum value in the sample.

<Stepwise>

Stepwise removes and adds terms to the model for the purpose of identifying a useful subset of the terms. Minitab provides three commonly used procedures: standard stepwise (adds and removes terms), forward selection (adds terms), and backward elimination (removes terms).

If you choose a stepwise procedure, the terms that you specify in the **Model** dialog are candidates for the final model.

See Using automatic selection procedures for a discussion of potential problems with stepwise procedures.

DIALOG BOX ITEMS

Method: Specify the method that Minitab uses to fit the model.

None: Choose to fit the model with all of the terms that you specify in the **Model** dialog.

Stepwise: By default, this procedure starts with an empty model and then adds or removes a term for each step. You can specify terms to include in the initial model or to force into every model.

Forward selection: By default, this procedure starts with an empty model and adds the most significant term for each step. You can specify terms to force into every model.

Backward elimination: This procedure starts with all potential terms in the model and removes the least significant term for each step. You can specify terms to force into every model.

Potential terms: Displays the set of terms that the procedure will assess. Indicators in the list signify how the procedure handles the term. The **Method** you choose determines the initial settings in this list. You can modify how the procedure handles the terms with the two buttons below. The indicators are:

- **E:** The term will be included in every model regardless of its p-value.

- **I:** The term will be included in the initial model but it can be removed from the model by the procedure if its p-value is too high.

- **<blank>:** The procedure can add or remove the term from the model based on its p-value.

E = Include term in every model: Select a term and click this button to force the term into every model regardless of its p-value. Click the button again to remove this condition.

I = Include term in initial model: Select a term and click this button to include the term in the initial model. Click the button again to remove this condition. This button is only available if you choose **Stepwise** in **Method**.

Alpha to enter: Enter the criterion for entering a new term in the model. You can set this value when you choose **Stepwise** or **Forward selection**.

Alpha to remove: Enter the criterion for removing a term from the model. You can set this value when you choose **Stepwise** or **Backward elimination**.

Hierarchy: This dialog allows you to control how Minitab enforces model hierarchy during stepwise procedures. This button is disabled if you specify a non-hierarchical model or uncheck **Include the constant term in the model** in the **Model** dialog.

Display the table of model selection details: Check to display information about the stepwise procedure in the Session window.

Details about the method: Choose to display the type of stepwise procedure and the alpha values to enter and/or remove a predictor from the model.

Include details for each step: Choose to display the coefficients , p-values , Mallows' Cp , and model summary statistics for each step of the procedure.

<Graphs>

Stat > Regression > Regression > Fit Regression Model > Graphs

Display [residual plots](#) for diagnosis of the regression model fit.

Residuals for Plots: You can specify the type of residual to display on the residual plots. See [Choosing a residual type](#).

Regular: Choose to plot the regular or raw [residuals](#) .

Standardized: Choose to plot the [standardized residuals](#) .

Deleted: Choose to plot the [Studentized deleted residuals](#) .

Residual Plots

Individual plots: Choose to display one or more plots.

Histogram of residuals: Check to display a histogram of the residuals.

Normal probability plot of the residuals: Check to display a normal probability plot of the residuals.

Residuals versus fits: Check to plot the residuals versus the fitted values.

Residuals versus order: Check to plot the residuals versus the order of the data. The row number for each data point is shown on the x-axis. For example, 1 2 3 4... n.

Four in one: Choose to display a layout of a histogram of residuals, a normal plot of residuals, a plot of residuals versus fits, and a plot of residuals versus order.

Residuals versus the variables: Enter one or more columns containing the variables against which you want to plot the residuals. Minitab displays a separate graph for each column.

[<Results>](#)

Stat > Regression > Regression > Fit Regression Model > Results

Control the level of detail in the display of output to the Session window.

Display of results: Choose the level of detail to display.

Simple tables: Choose to display the simpler versions of all tables.

Expanded tables: Choose to display the expanded versions of the ANOVA, Coefficients, Model summary, and Fits and Diagnostics tables.

Method: Check to display a table that summarizes non-default settings.

Analysis of variance: Check to display the [ANOVA table](#) that includes the sums of squares by source and the p-values.

Model summary: Check to display statistics that evaluate model fit, including [R^2](#) .

Coefficients: Check to display the [coefficients](#) , effects, and p-values. If your model contains categorical predictors, the drop-down list is enabled so that you can control how many coefficients for categorical predictors are in the table.

Default coefficients: Choose to display all linearly independent coefficients of the categorical predictors.

Full set of coefficients: Choose to display coefficients for all levels of the categorical predictors, which includes the final, linearly dependent, level.

Regression equation: Check to display the regression equation(s). If your model contains categorical predictors, the drop-down list is enabled so that you can control how many equations are displayed.

Separate equation for each combination of categorical predictor levels: Choose to display a separate equation for each combination of categorical predictors.

Single equation: Choose to display one equation that includes all levels of all categorical predictors.

Fits and diagnostics: Check to display fits and diagnostics.

Only for unusual observations: Choose to display fits and diagnostics for only the [unusual observations](#) .

For all observations: Choose to display fits and diagnostics for all observations.

Durbin-Watson statistic: Check to display the [Durbin-Watson](#) statistic to test for autocorrelation.

[<Storage>](#)

Stat > Regression > Regression > Fit Regression Model > Storage

You can store diagnostic measures and characteristics of the [estimated regression equation](#) for future use. Minitab stores the checked values in the next available columns and names the columns.

DIALOG BOX ITEMS

Fits: Check to store the [fitted values](#) .

Residuals: Check to store the [residuals](#) .

Standardized residuals: Check to store the [standardized residuals](#) .

Deleted residuals: Check to store the [Studentized deleted residuals](#) .

Leverages: Check to store the [leverages](#) .

Cook's distance: Check to store [Cook's distance](#) .

DFITS: Check to store the [DFITS](#) .

Coefficients: Check to store the [estimated coefficients](#) of the regression equation. These are the same coefficients as are printed in the output. If some terms are removed because the data cannot support them, the removed terms do not appear on the output.

Design matrix: Check to store the [design matrix](#) that corresponds to your model. Minitab takes the factors, creates the squares and cross-products, and stores all of these in a matrix. To view the matrix in the Session window, choose **Data > Display Data**. This stores the design matrix for use in other Minitab commands. To view the matrix in the worksheet, choose **Data > Copy > Matrix to Columns**.

Box-Cox Transformation: If you transform your response variable, you can store information about the transformation.

Box-Cox transformation of response: Check to store the transformed values of the response variable.

Fits for original response: Check to store the fits for the original response.

MINITAB ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance (ANOVA) is similar to regression in that it is used to investigate and model the relationship between a response variable and one or more predictor variables . However, analysis of variance differs from regression in two ways: the predictor variables tend to be categorical . In effect, analysis of variance extends the two-sample t-test for testing the equality of two population means to a more general null hypothesis of comparing the equality of more than two means, versus them not all being equal. Several of Minitab's ANOVA procedures, however, allow models with both categorical and continuous variables.

ONE-WAY ANOVA

The Stat > Anova > One-way option is used to determine whether the means of two or more groups differ, obtain a range of values for the difference between each pair of groups, and graph your data. Minitab performs the traditional One-Way ANOVA procedure if you assume that the variance is constant across all groups. However, if your data exhibit unequal variances between groups, Minitab can perform Welch's ANOVA. To perform Welch's ANOVA, uncheck Assume equal variances in the Options dialog.

DIALOG BOX ITEMS

In the dropbox list box, choose the option that matches your data as *Response data are in one column for all factor levels* or *Response data are in a separate column for each factor level*.

Choose the Response data are in one column for all factor levels option when data are in one column for all factors. Enter the continuous measurements in the Response: textbox. Enter which group each measurement belongs to in the Factor: textbox.

Choose the Response data are in a separate column for each factor level option when data are in a separate column for each factor level. In the Responses: textbox, enter the columns that contain the values for the response. Each column contains the continuous measurements for a separate group.

Consider three alternatives to assembling a new filtration system. The response variable is the number of filtration systems assembled per week.
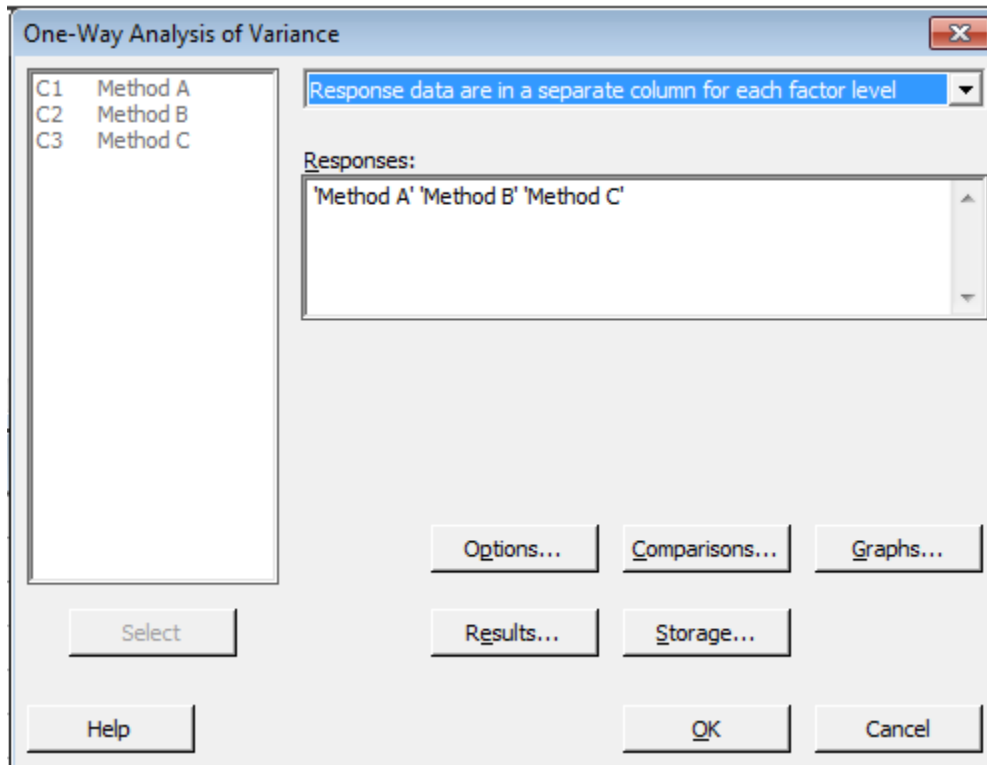
In the One-Way Analysis of Variance: Options dialog box, check the Assume equal variances: checkbox to assume that all populations have equal variances. If you do not assume equal variances, Minitab performs Welch's ANOVA. One-way ANOVA with equal variances is slightly more powerful than Welch's ANOVA, but serious error can result if the variances are not equal. Enter the confidence level in the Confidence level: textbox as a number between 0 and 100. The default is 95. Choose the Type of confidence interval: as a two-sided, lower bound, or upper bound confidence interval.
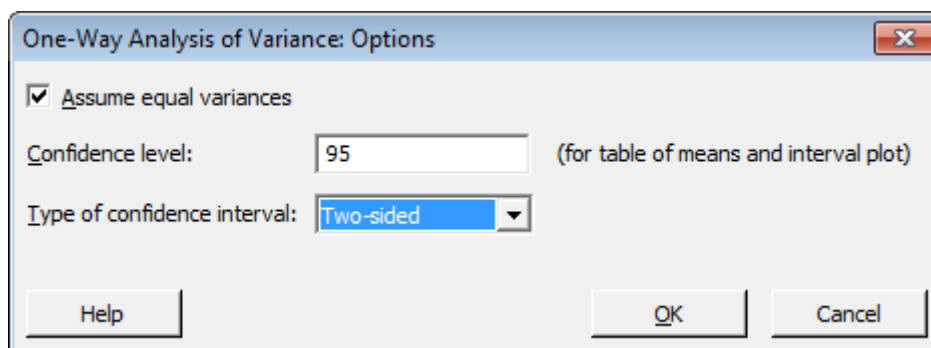
In the One-Way Analysis of Variance: Comparisons options box, choose the comparison procedure based on the group means that you want to compare and the error rate that you want to specify. Procedures that do not compare all pairs of groups are more powerful. The choice of comparison method depends on your desired inference. Using Tukey's all-differences approach is inefficient when Dunnett's or Hsu's MCB is suitable, because Tukey's confidence intervals are wider and the hypothesis tests less powerful for a given family error rate. For the same reasons, Hsu's MCB is superior to Dunnett's if you want to eliminate levels that are not the best and to identify those that are best or close to the best. The choice between Tukey's and Fisher's depends on whether you want to specify the family or individual error rate .

DIALOG BOX ITEMS

Enter an error rate in the Error rate for comparisons: textbox as a number between 0.001 and 50. The default is 5. The error rate is the individual error rate for Fisher's method and the family error rate for all other methods.

Choose the Comparison procedures assuming equal variances as Tukey, Fisher, Dunnett, or HSU's MCB. Check the Tukey: checkbox to compare all pairs of groups, while controlling the family error rate. Tukey's provides confidence intervals that are less precise and hypothesis tests that are less powerful than either Dunnett's or Hsu's MCB. Check the Fisher: checkbox to compare all pairs of groups, while controlling the individual error rate. Fisher's is less common than Tukey's because it does not control the family error rate, which often increases to unacceptable levels. Check the Dunnett: checkbox to compare the treatment groups to a control group, while controlling the family error rate. Choose the value for the control group level in the Control group level: dropdown list box. Check the Hsu MCB: checkbox to identify the groups that are the best. Hsu MCB compares each group to the group with either the largest or the smallest mean, while controlling the family error rate. In the Best: dropdown list box, choose Largest mean is best to compare each group to the group with the largest mean. Choose Smallest mean is best to compare each group to the group with the smallest mean.

In the Results section, check the Interval plot for differences of means: checkbox to display confidence intervals for the differences between the group means for each selected comparison method. Use the confidence intervals to determine likely ranges for the differences and to assess the practical significance of the differences. Check the Grouping information: checkbox to display a table that highlights the significant and non-significant comparisons for each selected comparison method. Check the Tests: checkbox to display the hypothesis test form of the comparison output, which includes the differences of the means and the adjusted p-values .

In the One-Way Analysis of Variance: Graphs dialog box, choose to display an interval plot, an individual value plot, a boxplot, and residual plots. If the response data are in separate columns for each factor level, the Residuals versus order and the Residuals versus the variables plots are not available.
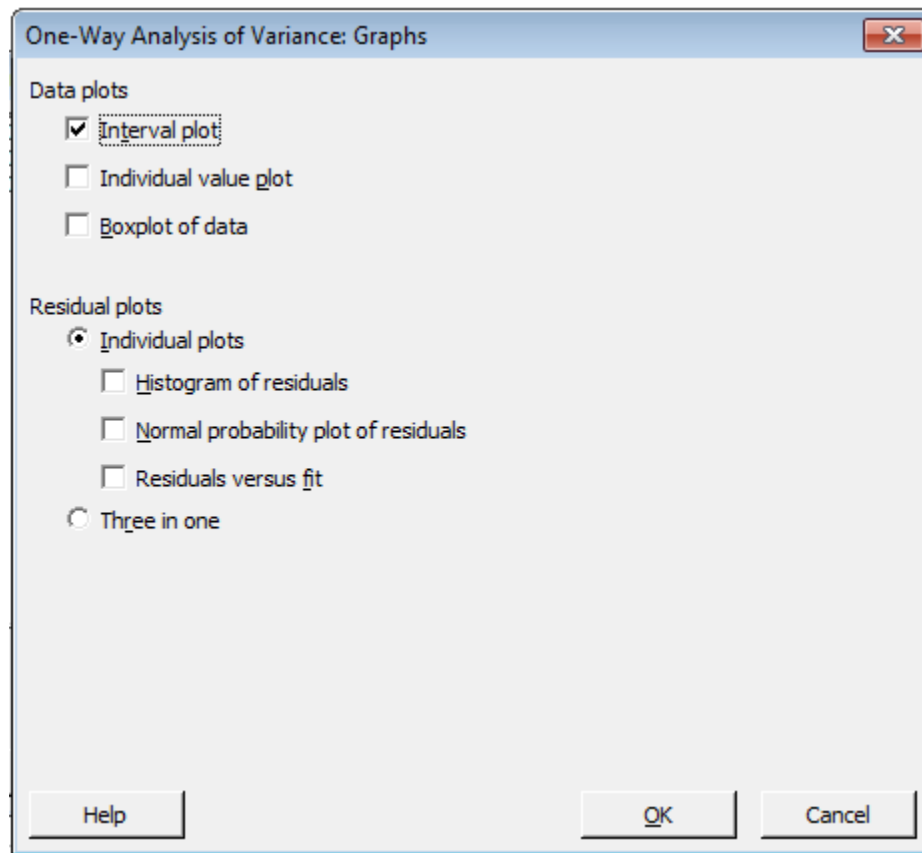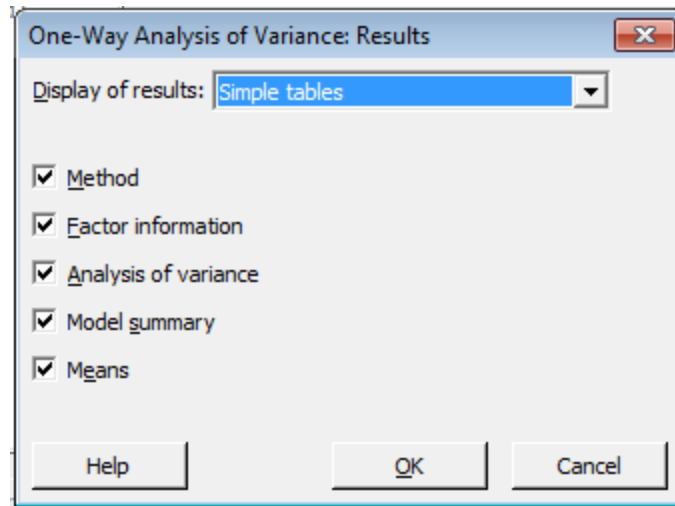
DIALOG BOX ITEMS

Check the Interval plot: checkbox to display an interval plot of each sample. Check the Individual value plot: checkbox to display an individual value plot of each sample. Check the Boxplots of data: checkbox to display a boxplot of each sample.

In the Residual plots section, choose Individual plots: to display one or more plots. Check the Histogram of residuals: checkbox to display a histogram of the residuals. Check the Normal probability plot of residuals: checkbox to display a normal probability plot of the residuals. Check the Residuals versus fits: checkbox to plot the residuals versus the fitted values.

In the One-Way Analysis of Variance: Results dialog box, control the level of detail in the display of output to the Session window.

In the Display of results: dropdown list box, choose the level of detail to display as *Simple tables* or *Expanded tables*. Choose *Simple tables* to display the simpler versions of all tables. Choose *Expanded tables* to display the expanded versions of the ANOVA table, Model summary, and Multiple comparisons tables. Check the Method: checkbox to display a table that summarizes the settings. Check the Factor information: checkbox to display the factor name, number of levels, and level values. Check the Analysis of variance: checkbox to display the ANOVA table or Welch's test table that includes the p-values. Check the Model summary: checkbox to display statistics that evaluate model fit, including R2. Check the Means: checkbox to display the factor level means, standard deviations, and confidence intervals.

In the One-Way Analysis of Variance: Storage dialog box, store diagnostic measures for future use. Minitab stores the checked values in the next available columns and names the columns.
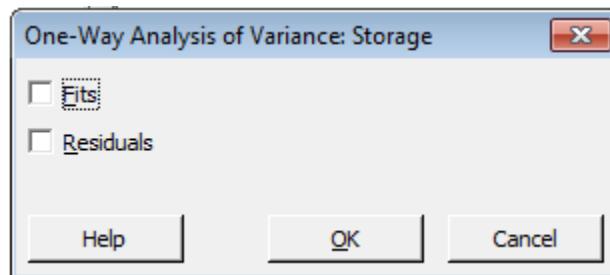
DIALOG BOX ITEMS

Check the Fits: checkbox to store the fitted values for the group means. Check the Residuals: checkbox to store the residuals.

One-way ANOVA: Method A, Method B, Method C

Method

```
Null hypothesis          All means are equal
Alternative hypothesis   At least one mean is different
Significance level       α = 0.05
```

Equal variances were assumed for the analysis.

Factor Information

```
Factor   Levels  Values
Factor        3  Method A, Method B, Method C
```

Analysis of Variance

```
Source  DF  Adj SS  Adj MS  F-Value  P-Value
Factor   2   520.0  260.00     9.18    0.004
Error   12   340.0   28.33
Total   14   860.0
```

Model Summary

```
      S   R-sq  R-sq(adj)  R-sq(pred)
5.32291  60.47%    53.88%      38.23%
```

Means

```
Factor     N   Mean  StDev       95% CI
Method A   5  62.00   5.24  (56.81, 67.19)
Method B   5  66.00   5.15  (60.81, 71.19)
Method C   5  52.00   5.57  (46.81, 57.19)
```

Pooled StDev = 5.32291

CHI-SQUARE GOODNESS OF FIT TEST

Stat > Tables > Chi-Square Goodness-of-Fit Test (One Variable)

Use Chi-Square Goodness-of-Fit Test to test the hypotheses:

H_0: Data follow a multinomial distribution with certain proportions

H_1: Data do not follow a multinomial distribution with certain proportions

The results may not be accurate if the expected frequency of any category is less than 5.

DIALOG BOX ITEMS

Observed counts: Choose if you have summary values of observed counts for each category. Enter the column containing the observed counts or type the observed counts for each category.

Category names (optional): Enter the column containing the category names or type each category's name.

Categorical data: Choose if you have raw categorical data in a column. Enter the column.

Test

Equal proportions: Choose to assume equal proportions across categories.

Specific proportions: Choose if you have different proportions for each category. If you have proportions for each category stored in a column, choose **Input column** and enter the column. If you want to type the proportion for each category, choose **Input constants** and type the proportions for the corresponding categories.

Proportions specified by historical counts: Choose if you have historical counts for each category. If you have historical counts for each category stored in a column, choose **Input column** and enter the column. If you want to type the historical counts for each category, choose **Input constants** and type the historical count for the corresponding categories.

Stat > Tables > Chi-Square Goodness-of-Fit Test (One Variable) > Graphs

Displays the bar chart of observed and expected values and each category's contribution to the chi-square value .

Bar chart of the observed and the expected values: Choose to display a bar chart of observed and expected values.

Bar chart of each category's contribution to the Chi-Square value: Choose to display a bar chart of each category's contribution to the chi-square value.

Display bars from the largest to the smallest: Choose to display each category's contribution to the chi-square value with bars ordered from largest to smallest.

Stat > Tables > Chi-Square Goodness-of-Fit Test (One Variable) > Results

You can control the display of test results in the Session window.

Display test results: Choose to display observed and expected counts and the test results.

Stat > Tables > Chi-Square Test for Association

Perform a chi-square test of association between variables. The null hypothesis states that no association exists.

Raw data (categorical variables)

Rows: Enter the column containing the categories that define the rows of the table.

Columns: Enter the column containing the categories that define the columns of the table.

Summarized data in a two-way table

Columns containing the table: Enter the columns containing the contingency table data. Delete or combine rows or columns with missing data before using this procedure.

Labels for the table (optional): Use these options to add labels to the display of your output.

Rows: (Optional) Enter a column that contains labels for the values of the row variable. The name of this column becomes the name of the row variable.

Columns: (Optional) Enter the name of the column variable.

Stat > Tables > Chi-Square Test for Association > Statistics

You can perform tests of independence and control the display of statistics for measures of association.

DIALOG BOX ITEMS

Chi-square test: Check to perform a chi-square test for association.

Counts

Display counts in each cell: Check to display the observed count for each combination of row and column variables.

Display marginal counts: Check to display the total of the observed counts for each row and column of the table.

Statistics to display in each cell

Expected cell counts: Check to display each cell's expected count.

Raw residuals: Check to display the difference between the observed and the expected count in each cell.

Standardized residuals: Check to display the standardized residuals of each cell.

Adjusted residuals: Check to display the adjusted residuals of each cell.

Each cell's contribution to chi-square: Check to display each cell's contribution to the overall chi-square statistic . The contribution to the chi-square statistic is the (standardized residual) for each cell.

Stat > Tables > Chi-Square Test for Association > Options

You can include missing data in the output table and calculations.

Note   These options are not available for summarized data.

Missing values

Display for rows: Check to include a category in the table for missing values of the Row variable.

Display for columns: Check to include a category in the table for missing values of the Column variable.

Include displayed missing values in calculations Check to include missing values in calculations. Minitab uses the displayed missing values that you selected above in calculations. By default, Minitab does not include missing values in calculations.

GRAPHING DATA

Before you perform a statistical analysis, you can use graphs to explore data and assess relationships between the variables. Graphs can also be used to summarize data and to help interpret statistical results. Minitab's graphs may be accessed from the **Graph** and **Stat** menus.

HISTOGRAM

A histogram can be used to assess the shape and spread of continuous sample data. The histogram might be used to determine determine whether data follow a normal distribution. Minitab divides sample values into many intervals called bins. Bars (by default) represent the number of observations falling within each bin (its frequency).

From the Graph menu, choose Histogram. Four options are available for selection: *Simple*, *With Fit*, *With Groups*, and *With Fit and Groups*. Choose the *Simple* option to create a graph only. Select OK to continue.
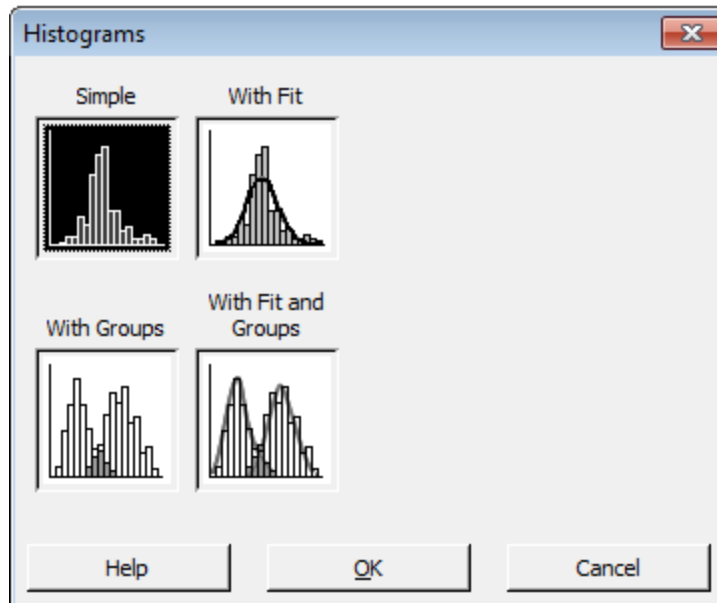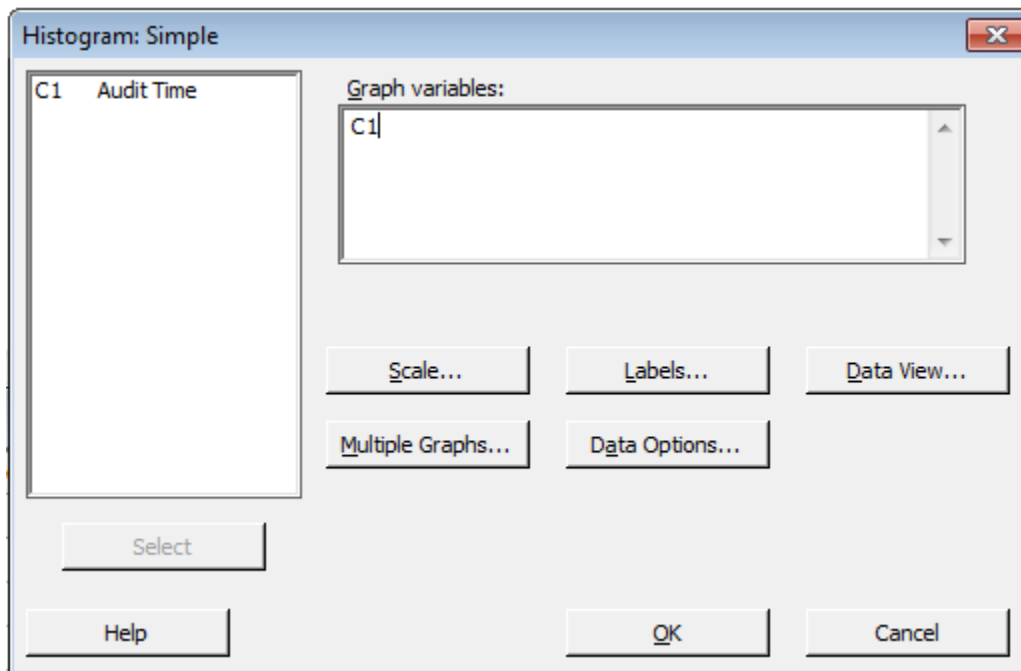
In the **Graph variables:** box enter either the column address or variable name. The Scale…, Labels…, Data View…, Multiple Graphs…, and Data Options… command buttons provide options for customizing the histogram presentation:

Scale…   Provide axes and ticks, indicate the Y scale type, add Gridlines, add References lines

Labels…   Add Title/Footnotes to the chart, add Data Labels

Data View…   Indicate the Data Display, Lowess, and Distribution of the data

Multiple Graphs…   Indicate Multiple variables, By variables, or By variables (area graphs, time series plots)

Data Options…   Indicate Subset, Group Options, Frequency
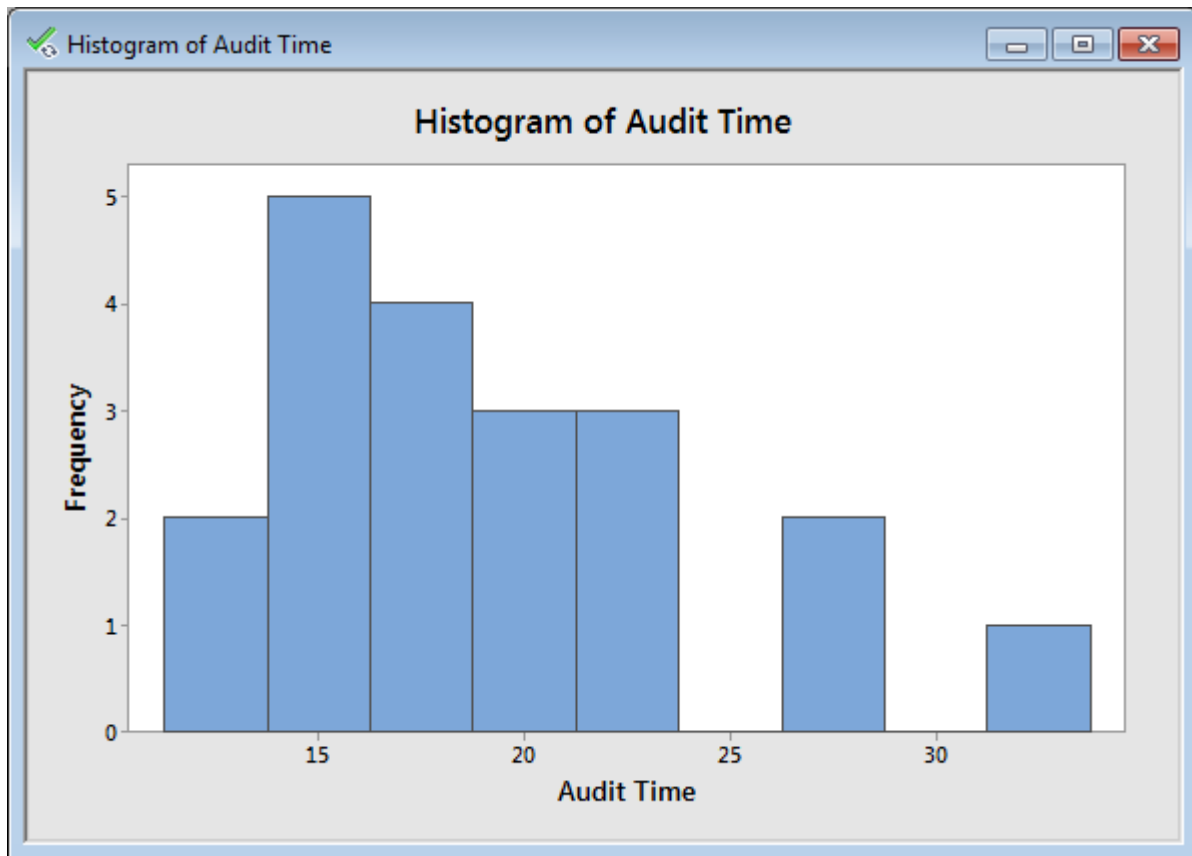
Position the mouse pointer over any one of the bars and Double-click. The Edit Bars dialog box appears.

EDIT BARS DIALOG BOX
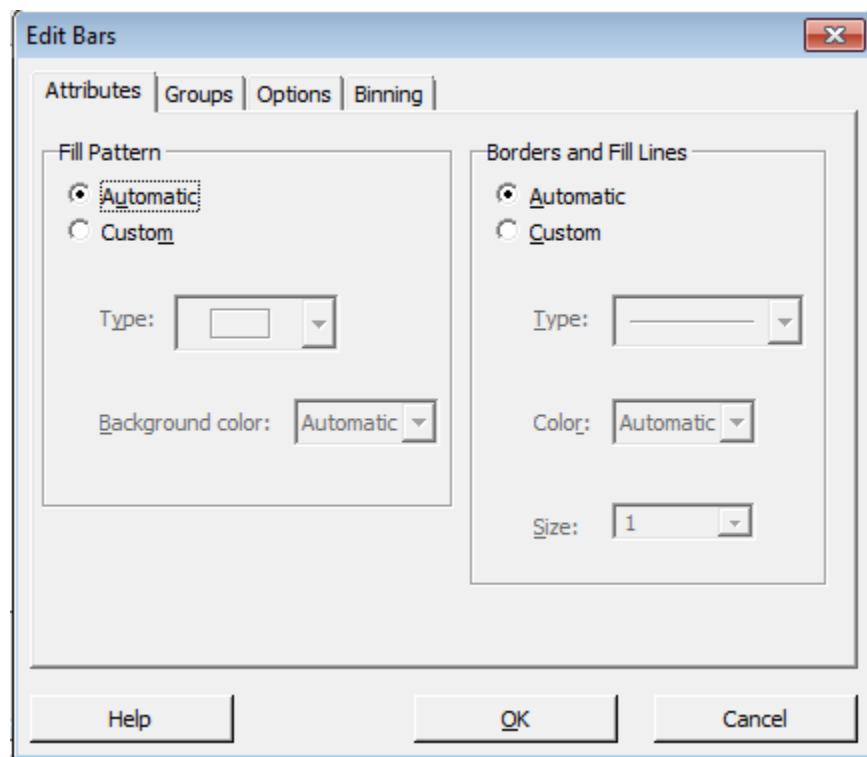
BINNING TAB

In the Binning tab, select Cutpoint for Interval Type. Choose the Midpoint/Cutpoint positions: and indicate 10:35/5. The resulting histogram indicates a starting audit time value of 10, ending audit value of 35, and constructs histogram bars based on intervals of 5.
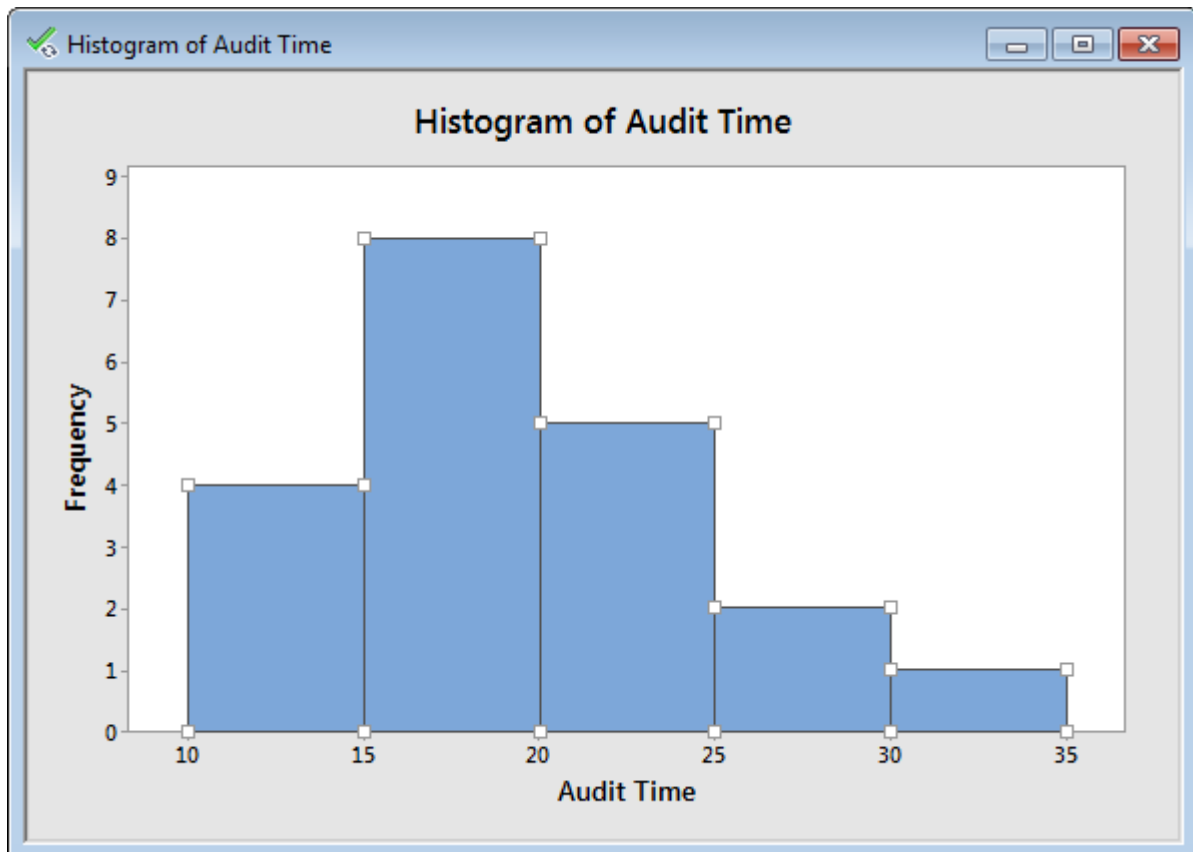
Use a dot plot to assess and compare distributions by plotting the values along a number line. Dot plots are especially useful for comparing distributions. The x-axis for a dot plot is divided into many small intervals, or bins. Data values falling within each bin are represented by dots. If possible, Minitab displays a dot for each observation. Otherwise, a dot represents multiple observations with a footnote indicating the maximum number of observations represented by each dot.

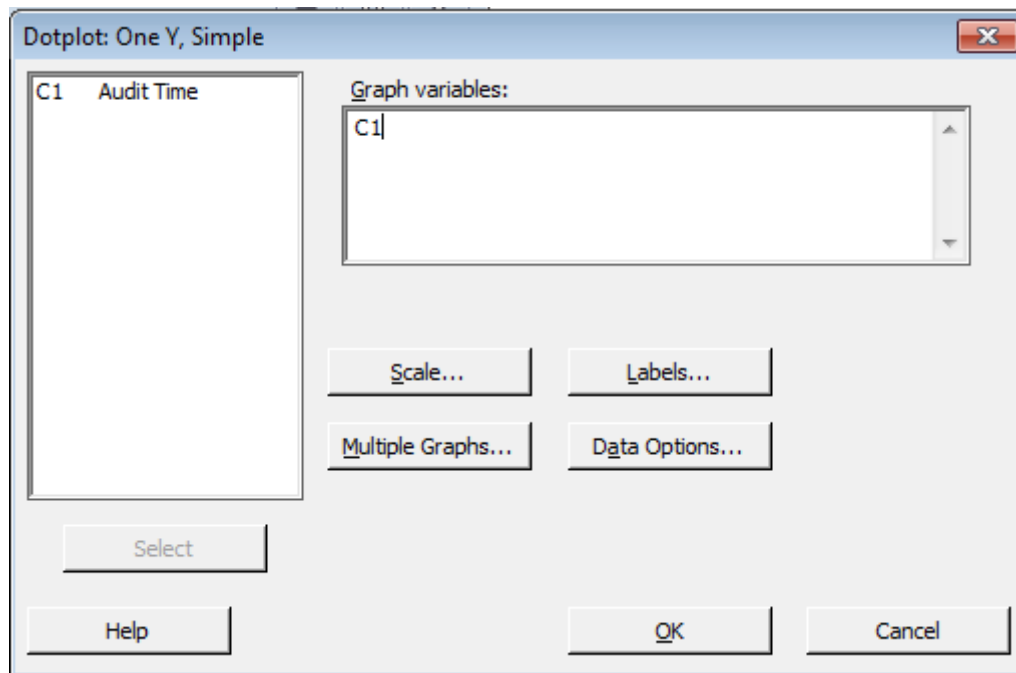From the Graph menu, choose DotPlot…
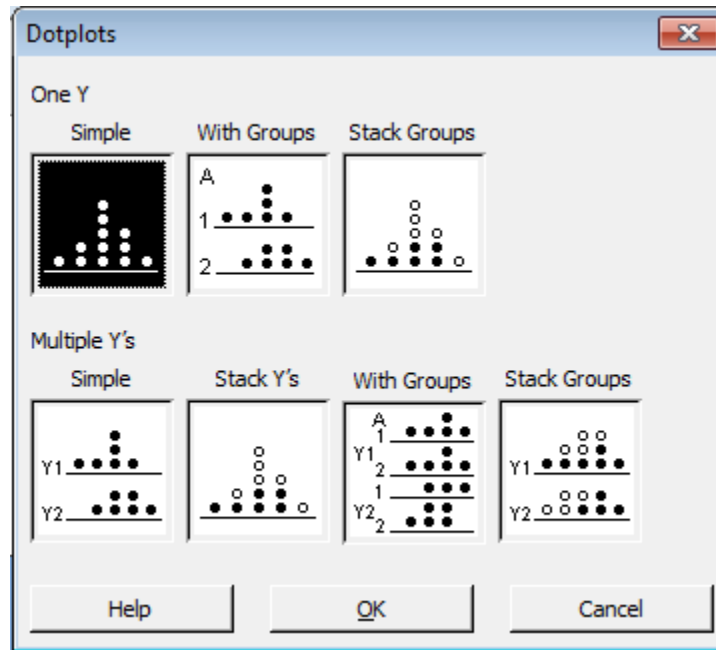
DOTPLOTS DIALOG BOX

One Y: Choose one of these options (Simple, With Groups, Stack Groups) to display a separate graph for each column.

Multiple Y's: Choose of of these options to display multiple columns of data on the same graphs.

For example, choose Simple and press OK

In the Dotplot: One Y, Simple dialog box indicate C1 as the variable to graph from the variables list. Press OK.
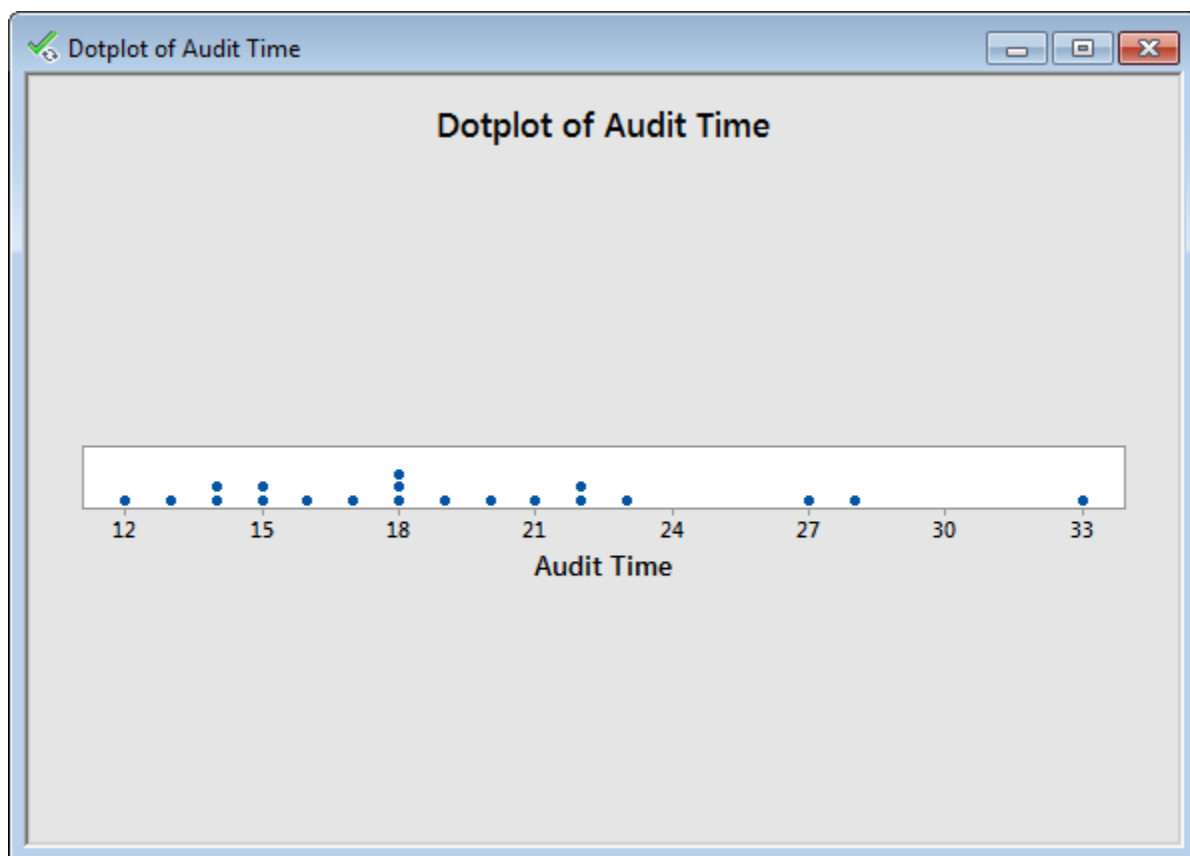
Use this tool (found in the Visualizing Data tab) to create 2- or 3–dimensional plots (represented by the size of the point).

Graph > Scatterplot

Use to illustrate the relationship between two variables by plotting one against the other.

Scatterplots are also useful for plotting a variable over time. Unlike a Time series plot , you must provide a time variable from the worksheet. This is especially useful for data that are not entered in chronological order or were collected at irregular intervals.

You need one or more pairs of equal-length columns of numeric or date/time data. You may also use up to three columns of categorical data for grouping. If either the x- or y-value for an observation is missing, Minitab does not plot the observation.

More For other data options, such as subsetting, using frequency columns, and handling missing values, see Data Options Overview. To change the display order for text categories, see Ordering text categories.

DIALOG BOX ITEMS

Simple	With Groups	With Regression	With Regression and Groups

With Connect Line	With Connect and Groups