StatCrunch™

Data analysis on the Web

www.statcrunch.com

PEARSON

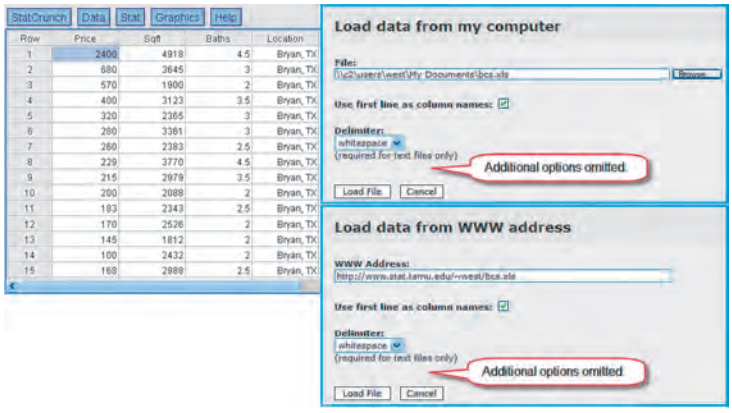
About This Study Card

StatCrunch is a Web-based statistical software package for analyzing data. To access StatCrunch, visit **http://www.statcrunch.com**. This study card is intended to serve as a brief introduction to the use of StatCrunch covering the procedures that most students will encounter in an introductory statistics course. Follow the help links at the statcrunch.com site for more extensive documentation.

Data > Load data > from file

1. Choose **My computer** to load a data file from the local system or **WWW** to load a data file from the Web.
2. Specify the location of a text file (.txt, .csv, etc.) or Microsoft Excel file.
3. If the first line in your file does not contain column names, deselect the **use first line as column names** option.
4. For text files (not Excel files), specify the delimiter for the data values. For example, the delimiter for a .csv file is a comma.
5. Click the **Load File** button to upload the data file.

About this data set: The data set contains information from a random sample of 30 four-bedroom homes listed for sale in the Bryan–College Station, Texas area in December 2008. For each home, the data set contains the list price in thousands of dollars (Price), square footage (Sqft), number of bathrooms (Baths), and location (Bryan, TX, or College Station, TX). It is currently being shared on the StatCrunch site at http://www.statcrunch.com/5.0/index.php?dataid=359673.

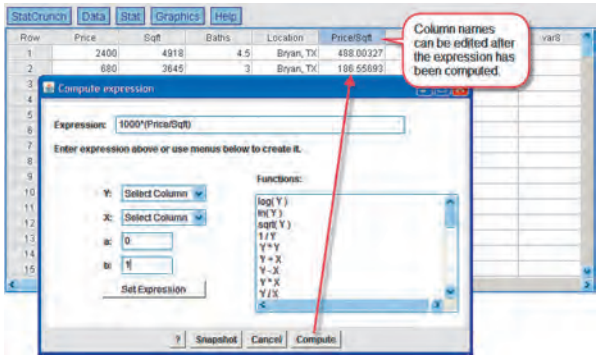


Example: Loading an Excel file from the local file system or from a Web address

Data > Compute Expression

1. Enter a mathematical or Boolean expression in the **Expression** input box.
2. Alternatively, use the menus for **Y, X, a, b,** and **Function** to construct the expression. Then, click **Set Expression**.
3. Click **Compute** and the results of the expression will be added as a new column to the StatCrunch data table.

Note: A Boolean (true/false) expression can be used as a **Where** statement in many StatCrunch procedures to exclude outliers or to focus an analysis on a subset of the data. Some of the following examples illustrate this feature.



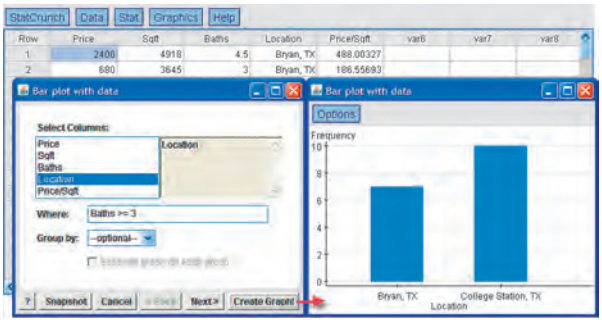
Example: Computing the mathematical expression, 1000*(Price/Sqft)

ISBN-13: 978-0-321-62892-3
ISBN-10: 0-321-62892-6
EAN 9 780321 628923 90000

Graphics > Bar plot

1. Choose the **with data** option to use data consisting of individual outcomes in the data table.
   a. Select the column(s) to be displayed.
   b. Use an optional **Where** statement to specify the data included.
   c. Select an optional **Group by** column for a side-by-side bar plot.

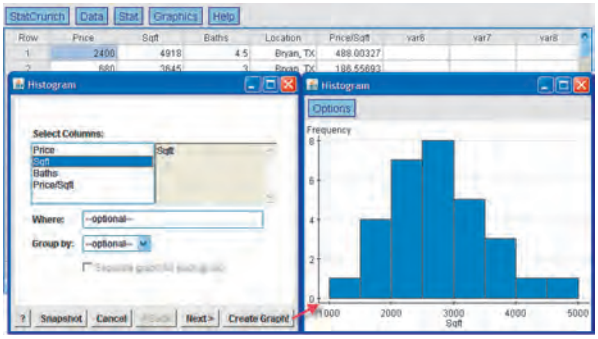
   Choose the **with summary** option to use summary information consisting of categories and counts.
   a. Select the column containing the categories.
   b. Select the column containing the counts.
2. Click **Next** to set additional options such as **Type** (Frequency or Relative Frequency).
3. Click **Create Graph!** to construct the bar plot(s).



Example: A bar plot showing the number of homes in each location with three or more baths

Graphics > Histogram

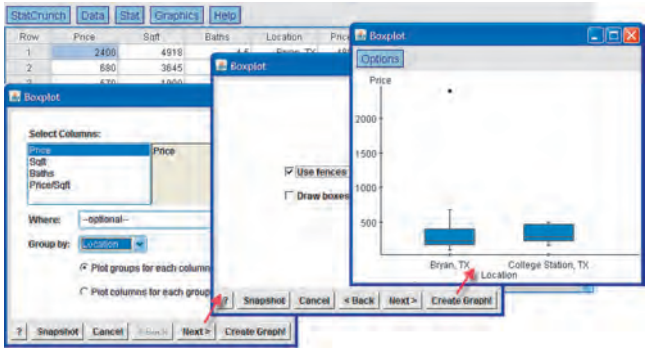
1. Select the column(s) to be displayed.
2. Use an optional **Where** statement to specify the data included.
3. Select an optional **Group by** column to color code the bars.
4. Click **Next** to set additional options such as the starting value of the bins and the bin width.
5. Click **Next** again to specify an optional density to overlay on the histogram(s).
6. Click **Create Graph!** to construct the histogram(s).



Example: A histogram of the Sqft column

Graphics > Boxplot

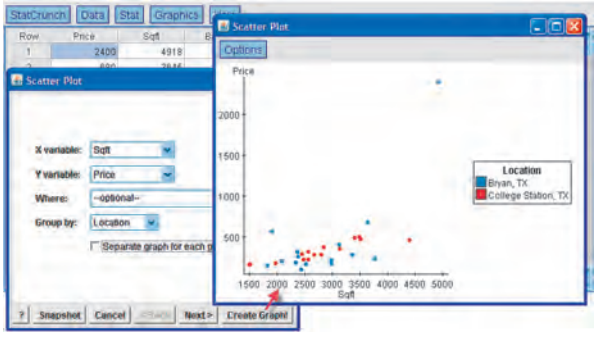
1. Select the column(s) to be displayed. By default, a boxplot for each column will be included on a single graph.
2. Use an optional **Where** statement to specify the data included.
3. Select an optional **Group by** column to compare boxplots across groups on a single graph.
4. Click **Next** to indicate whether or not to use fences for the boxplots. By default, the five-number summary is used.
5. Click **Create Graph!** to construct the boxplot(s).



Example: Boxplots comparing price across locations

Graphics > Scatter Plot

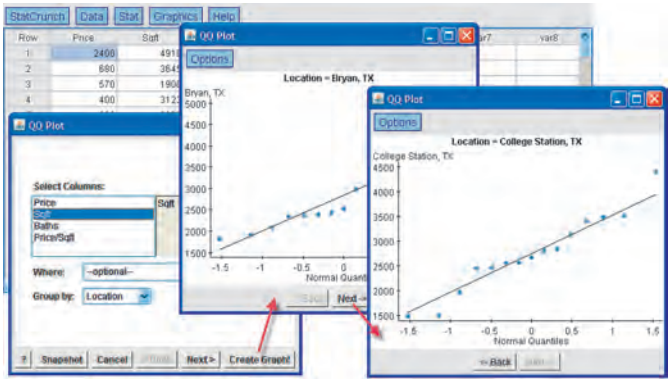
1. Select the **X variable** and **Y variable** for the plot.
2. Use an optional **Where** statement to specify the data included.
3. Color code points with an optional **Group by** column.
4. Click **Create Graph!** to produce the plot.



Example: Scatter plot of Price vs. Sqft color coded by location

Graphics > QQ Plot

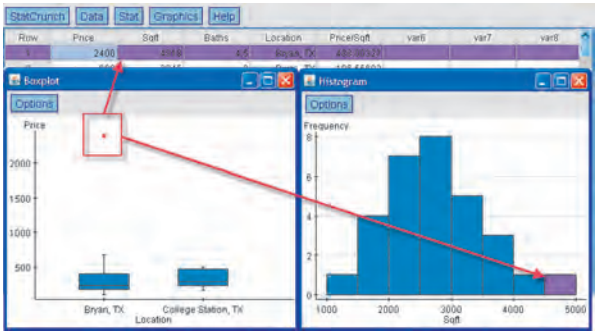
1. Select the column(s) to be displayed.
2. Use an optional **Where** statement to specify the data included.
3. Select an optional **Group by** column to produce separate plots for different groups.
4. Click **Create Graph!** to construct the plot(s).



Example: QQ plot of the Sqft column grouped by Location

Interacting with Graphics

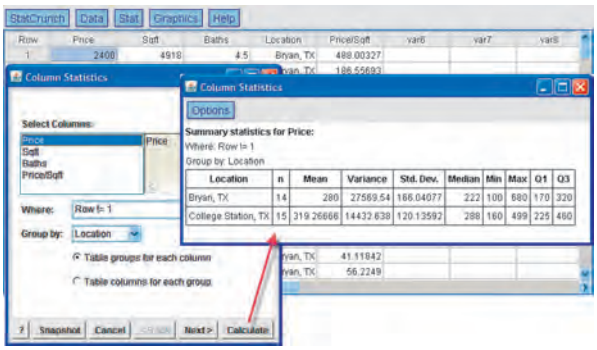
1. Click and drag the mouse around graph objects to highlight them.
2. The corresponding rows will be highlighted in the data table and in all other graphics.
3. Toggle highlighting on and off by clicking on the row number in the data table.
4. To clear all highlighted rows, click on the **Row** heading atop the first column in the data table.
5. To highlight rows based on categories or numeric ranges, use the **Data** > **Row selection** > **Interactive tools** option.



Example: Highlighting an outlier in a boxplot

Stat > Summary Stats > Columns

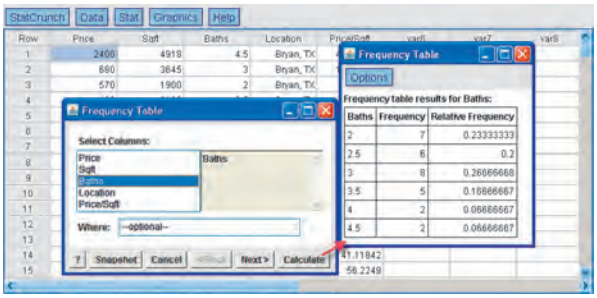
1. Select the column(s) for which summary statistics are to be computed.
2. Use an optional **Where** statement to specify the data included.
3. Compare statistics across groups using an optional **Group by** column.
4. Click **Next** for additional options such as the statistics to be computed.
5. Click **Calculate** to compute the summary statistics.



Example: Comparing prices of homes listed in Bryan to those listed in College Station with the potential outlier removed

Stat > Tables > Frequency

1. Select the column(s) for which a frequency table is to be computed.
2. Use an optional **Where** statement to specify the data included.
3. Click **Calculate** to compute the frequency table(s).

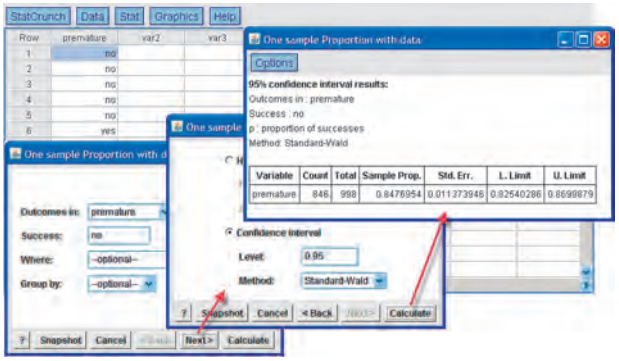


Example: A frequency table for the number of bathrooms

Stat > Proportions > One sample

1. Choose the **with data** option to use data from the data table.
   a. Select the column containing the sample values.
   b. Specify the outcome that denotes a **success**.
   c. Use an optional **Where** statement to specify the data included.

   Choose the **with summary** option to enter the **number of successes** and **number of observations**.
2. Click **Next** and select the **hypothesis test** or **confidence interval** option.
   a. For a hypothesis test, enter the **Null** proportion and choose $\neq$, $<$, or $>$ for **Alternative**.
   b. For a confidence interval, enter a value between 0 and 1 for **Level** (0.95 provides a 95% confidence interval). For **Method**, choose **Standard-Wald** or **Agresti-Coull**.
3. Click **Calculate** to view the results.

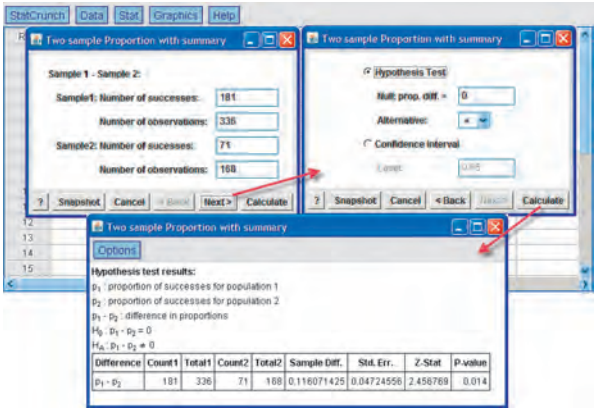


Example: For each of 998 North Carolina births, this data set indicates whether or not the birth was premature. A 95% confidence interval for the proportion of all North Carolina births that are *not* premature is shown above.

Stat > Proportions > Two sample

1. Choose the **with data** option to use sample data from the StatCrunch data table.
   a. Select the columns containing the first and second samples.
   b. Specify the sample outcomes that denote a success for both samples.
   c. Enter optional **Where** statements to specify the data rows to be included in both samples. If the two samples are in separate columns, this step is typically not required.

   Choose the **with summary** option to enter the **number of successes** and **number of observations** for both samples.
2. Click **Next** and select the **hypothesis test** or **confidence interval** option.
   a. For a hypothesis test, enter the **Null** proportion difference and choose $\neq$, $<$, or $>$ for **Alternative**.
   b. For a confidence interval, enter a value between 0 and 1 for **Level** (0.95 provides a 95% confidence interval).
3. Click **Calculate** to view the results.

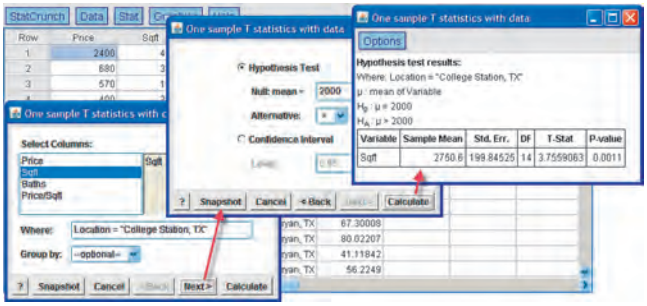


Example: In a survey, 181 of 336 people who attended church services at least once a week said the use of torture against suspected terrorists is "often" or "sometimes" justified. Only 71 of 168 people who "seldom or never" went to services agreed. Using this **summary** information, a two-sided hypothesis test is performed below to compare the proportion with this opinion for these two populations.

Stat > T statistics > One sample

1. Choose the **with data** option to use sample data from the StatCrunch data table.
   a. Select the column containing the sample data values.
   b. Use an optional **Where** statement to specify the data included.

   Choose the **with summary** option to enter the **sample mean**, **sample standard deviation**, and **sample size**.
2. Click **Next** and select the **hypothesis test** or **confidence interval** option.
   a. For a hypothesis test, enter the **Null** mean and choose $\neq$, $<$, or $>$ for **Alternative**.
   b. For a confidence interval, enter a value between 0 and 1 for **Level** (0.95 provides a 95% confidence interval).
3. Click **Calculate** to view the results.

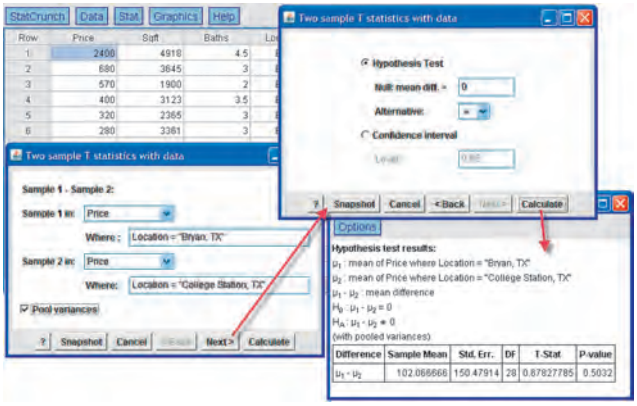


Example: Testing to see if the average home listed in College Station, Texas, is larger than 2000 square feet.

Stat > T statistics > Two sample

1. Choose the **with data** option to use sample data from the data table.
   a. Select the columns containing the first and second samples.
   b. Enter optional **Where** statements to specify the data rows to be included in both samples. If the two samples are in separate columns, this step is typically not required.

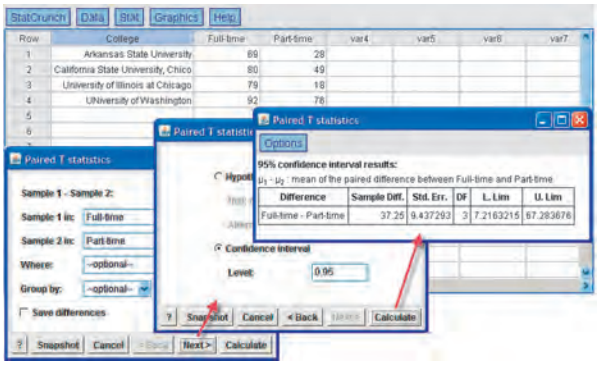
   Choose the **with summary** option to enter the **sample mean**, **sample standard deviation**, and **sample size** for both samples.
2. Deselect the **Pool variance** option if desired.
3. Click **Next** and select the **hypothesis test** or **confidence interval** option.
   a. For a hypothesis test, enter the **Null** mean difference and choose $\neq$, $<$, or $>$ for **Alternative**.
   b. For a confidence interval, enter a value between 0 and 1 for **Level** (0.95 provides a 95% confidence interval).
4. Click **Calculate** to view the results.



Example: Hypothesis test comparing the prices of homes listed in Bryan to those in College Station, with potential outliers included.

Stat > T statistics > Paired

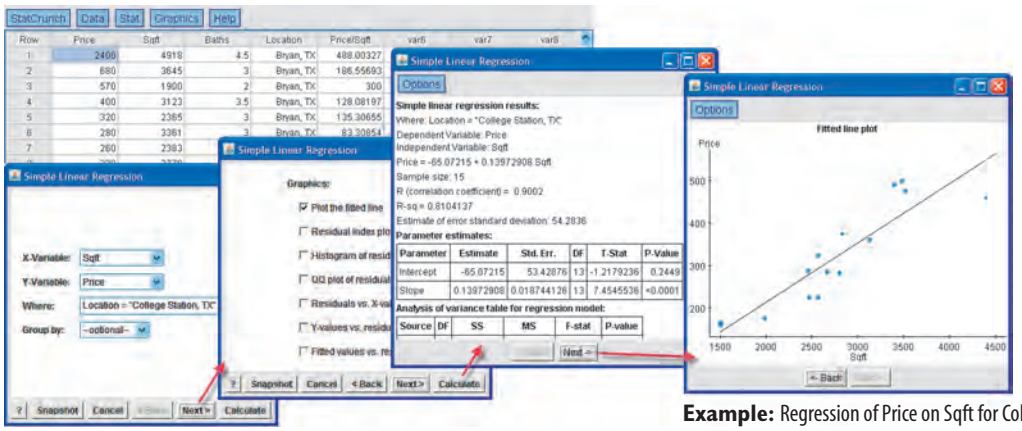
1. Select the columns containing the first and second samples.
2. Use an optional **Where** statement to specify the data included.
3. Click **Next** and select the **hypothesis test** or **confidence interval** option.
   a. For a hypothesis test, enter the **Null** mean difference and choose $\neq$, $<$, or $>$ for **Alternative**.
   b. For a confidence interval, enter a value between 0 and 1 for **Level** (0.95 provides a 95% confidence interval).
4. Click **Calculate** to view the results.



Example: A 95% confidence interval for the average difference between the graduation rate of full- and part-time students across all colleges based on a random sample of four colleges.

Stat > Regression > Simple Linear

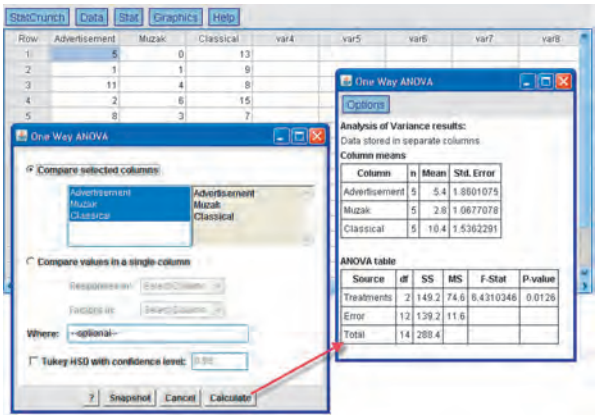
1. Select the **X variable** (independent variable) and **Y variable** (dependent variable) for the regression.
2. Use an optional **Where** statement to specify the data included.
3. Compare results across groups by selecting an optional **Group by** column.
4. Click **Next** to specify optional X values for predictions and to indicate whether or not to save residuals/fitted values.
5. Click **Next** again to choose from a list of optional graphics for plotting the fitted line and examining residuals.
6. Click **Calculate** to view the results.



Example: Regression of Price on Sqft for College Station

Stat > ANOVA > One way

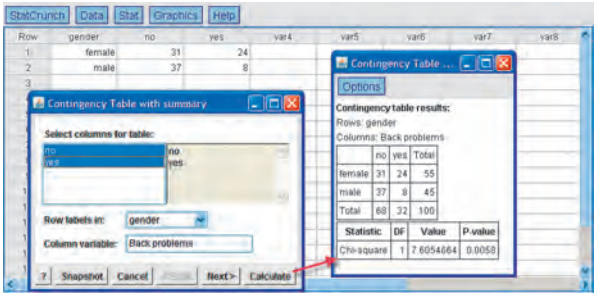
1. Select one of the following options:
   a. If the samples are in separate columns, select the **Compare selected columns** option and then select the columns containing the samples.
   b. If the samples are in a single column, select the **Compare values in a single column** option. Then specify the column containing the samples (**Responses In**) and the column containing the population labels (**Factors In**).
2. Use an optional **Where** statement to specify the data to be included.
3. Select the **Tukey HSD** option and specify a confidence level to perform a post hoc means analysis. The default value of 0.95 provides 95% confidence intervals for all pairwise mean differences.
4. Click **Calculate** to view the results.



Example: ANOVA for the time, in minutes, a caller stays on hold before hanging up under three different treatments.

Stat > Tables > Contingency

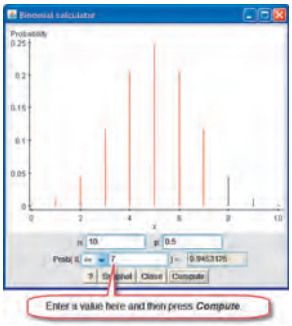
1. Select one of the following options:
   a. Choose the **with data** option to cross tabulate two columns of raw data from the data table. Then,
      A. Select the column to be tabulated across the rows.
      B. Select the column to be tabulated across the columns.
      C. Use an optional **Where** statement to specify the data included.
      D. Select an optional **Group by** column to compute separate tables across groups.
   b. Choose the **with summary** option to use a two-way cross classification already entered in the data table. Then,
      A. Select the columns that contain the summary counts.
      B. Select the column that contains the row labels.
      C. Enter a name for the column variable.
2. Click **Next** to specify additional information to be displayed in each table cell.
3. Click **Calculate** to view the results.



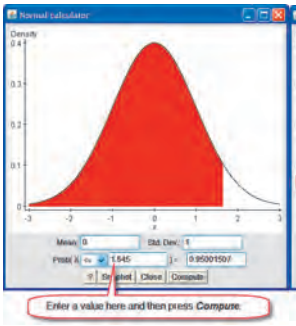
Example: Using **summary** counts cross tabulating gender and back problems, a test of independence between the two factors is shown below. Note that the back problems variable is represented by the second and third columns in the data table.

Stat > Calculators

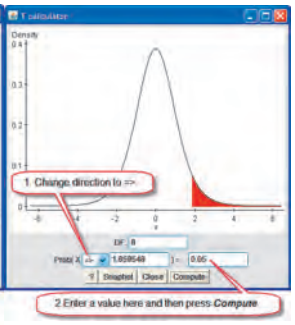
1. Select the name of the desired distribution from the menu listing (e.g., Binomial, Normal, etc.).
2. In the first line below the plot in the calculator window, specify the distribution parameters. As examples, with the normal distribution, specify the mean and standard deviation or with the binomial distribution, specify n and p.
3. In the second line below the plot, specify the direction of the desired probability.
   a. To compute a probability, enter a value to the right of the direction selector and leave the remaining field empty (e.g., $P(X < 3) =$ ___).
   b. To determine the point that will provide a specified probability, enter the probability to the right of the direction selector and leave the other field empty (e.g., $P(X >$ ___ $) = 0.25$). This option is available only for continuous distributions.
4. Click **Compute** to fill in the empty fields and to update the graph of the distribution.



Example: Finding a binomial probability



Example: Finding a standard normal probability



Example: Finding a t(8) quantile