



JMP Right In



Hello!

JMP (pronounced “jump”) software is so easy to use that after reading this chapter you’ll find yourself confident enough to learn everything on your own. Therefore, we cover the essentials fast—before you escape this book. This chapter offers you the opportunity to make a small investment in time for a large return later on.

If you are already familiar with JMP and want to dive right into statistics, you can skip ahead to Chapters 6–21. You can always return later for more details about using JMP or for more details about statistics.

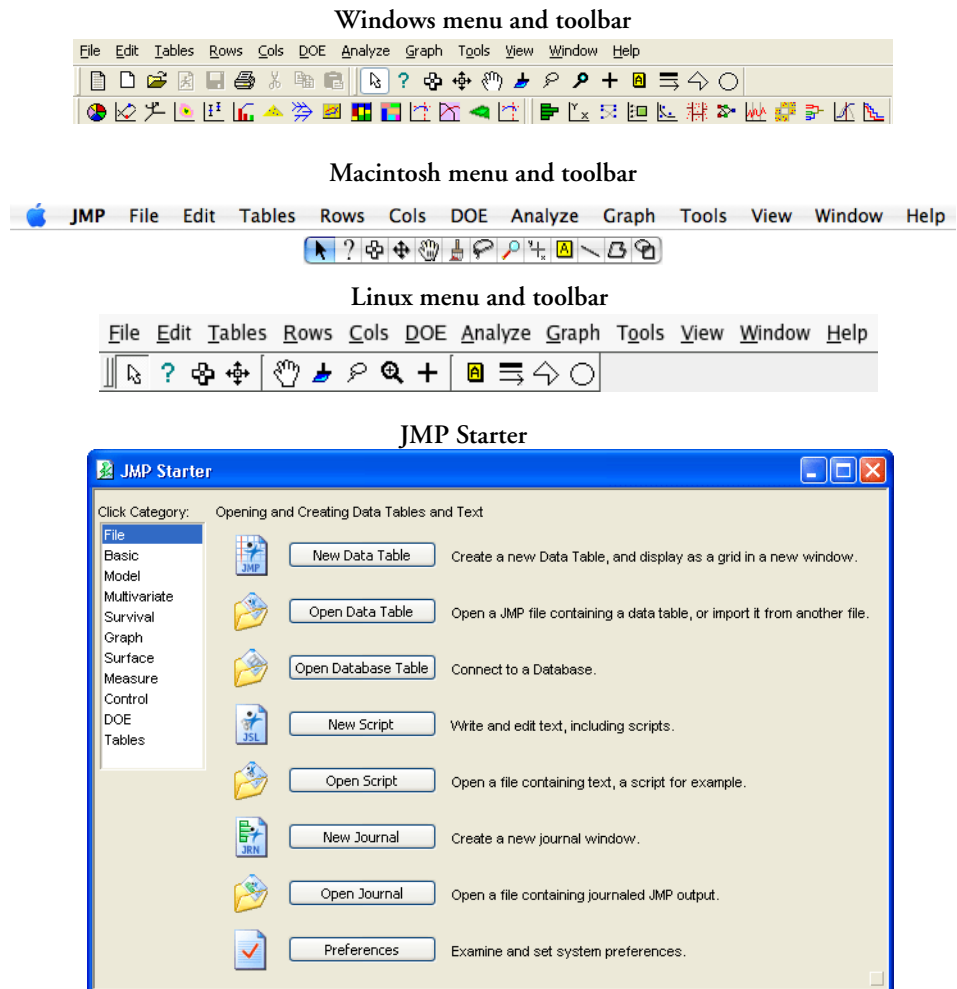
First Session

This first section just gets you started learning JMP. In most of the chapters of this book, you can follow along in a hands-on fashion. Watch for the mouse symbol (☞) and perform the action it describes. Try it now:

- ☞ To start JMP, double-click the JMP application icon.

When the application is active, you see the JMP menu bar and the JMP Starter window. You may also see toolbars, depending on how your system is set up. (Macintosh toolbars are attached to each window, and are appropriate for their window, and therefore vary.)

Figure 2.1 The JMP Main Menu and the JMP Starter



As with other applications, the **File** menu (**JMP** menu on Macintosh) has all the strategic commands, like opening data tables or saving them. To quit JMP, choose the **Exit** (Windows and Linux) or **Quit** (Macintosh) command from this menu. (Note that the **Quit** command is located on the **JMP** menu on the Macintosh.)

Start by opening a JMP data table and doing a simple analysis.

Open a JMP Data Table

When you first start JMP, you are presented with the JMP Starter window, a window that allows quick access to the most frequently used features of JMP. Instead of starting with a blank file or importing data from text files, open a JMP data table from the collection of sample data tables that comes with JMP.

- ☞ Choose the **Open** command in the **File** menu (choose **File > Open**).
- ☞ When the Open File dialog appears, as shown in **Figure 2.2**, **Figure 2.3**, or **Figure 2.4**, select **Big Class.jmp** from the list of sample data files.
- Windows sample data is usually installed at C:\Program Files\SAS\JMP7\English Support Files\Sample Data.
- Macintosh Sample Data is usually installed at the root level at /Library/Application Support/JMP/Support Files English/Sample Data.
- Linux Sample Data is usually installed at /JMP7/Support Files English/Sample Data in the directory where you installed JMP (typically /opt).
- ☞ Select **Big Class** and click **Open** (Windows and Macintosh) or **Finish** (Linux) on the dialog.

There is also a categorized list of the sample data, accessible from **Help > Sample Data Directory**. The pre-defined list of files may help you when searching through the samples. The above procedure was meant to show you how to, in general, open a data table.

Figure 2.2 Open File Dialog (Windows)

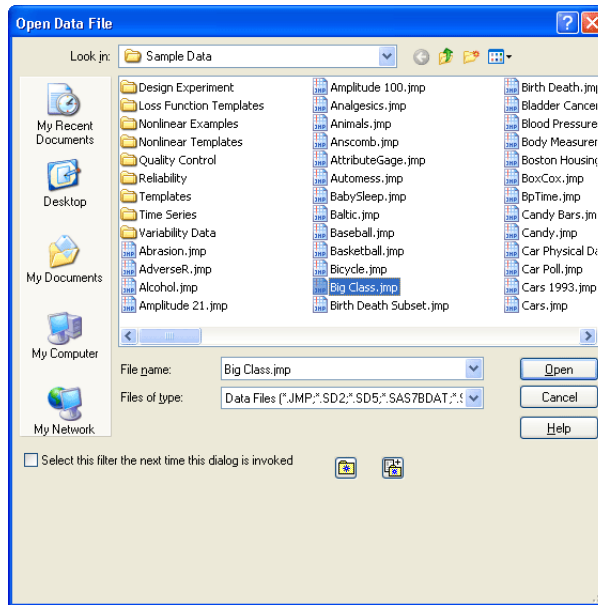


Figure 2.3 Open File Dialog (Macintosh)

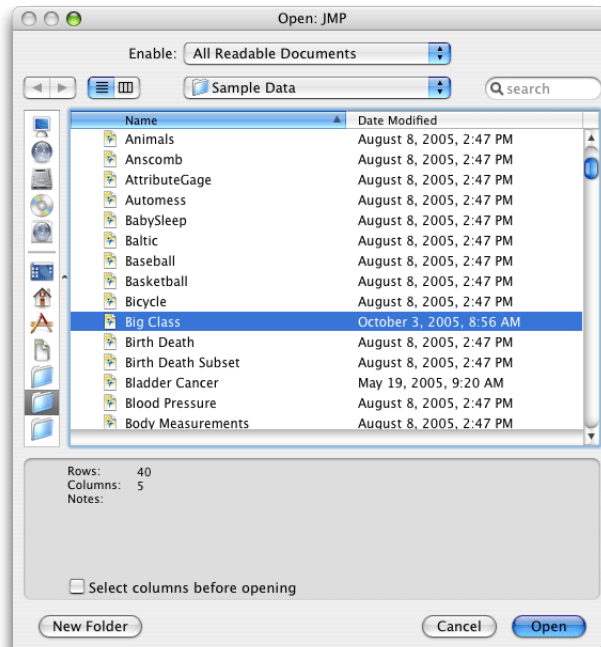
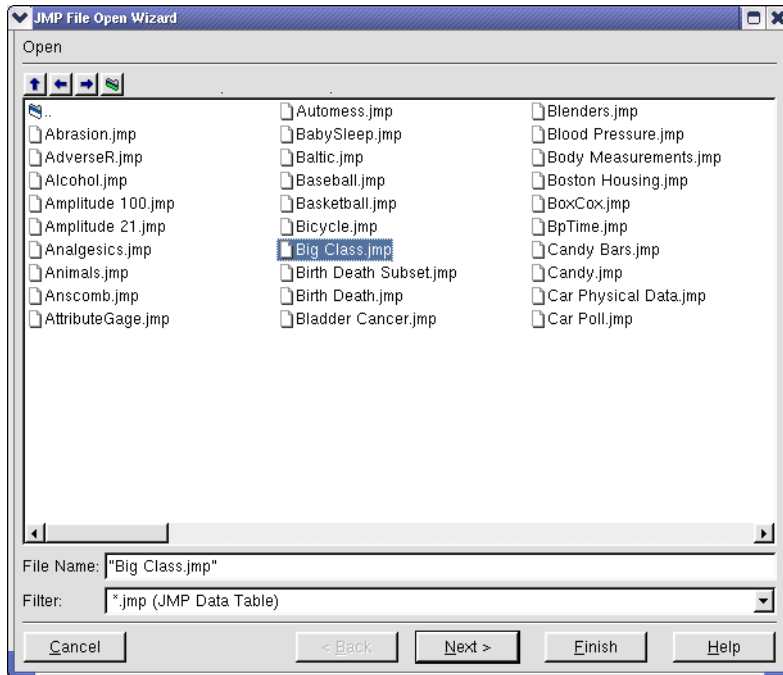


Figure 2.4 Open File Dialog (Linux)



You should now see a table with columns titled name, age, sex, height, and weight (shown in **Figure 2.5**).

In Chapter 3, “Data Tables, Reports, and Scripts” on page 27, you learn the details of the data table, but for now let’s try an analysis.

Figure 2.5 Partial Listing of the Big Class Data Table

▼ Big Class			name	age	sex	height	weight
▼ Distribution		1	KATIE	12	F	59	95
▼ Bivariate		2	LOUISE	12	F	61	123
▼ Oneway		3	JANE	12	F	55	74
▼ Logistic		4	JACLYN	12	F	66	145
▼ Contingency		5	LILLIE	12	F	52	64
▼ Fit Model		6	TIM	12	M	60	84
▼ Set Sex Value Labels		7	JAMES	12	M	61	128
▼ Set Age Value Labels		8	ROBERT	12	M	51	79
▼ Columns (5/0)		9	BARBARA	13	F	60	112
name		10	ALICE	13	F	61	107
age		11	SUSAN	13	F	56	67
sex		12	JOHN	13	M	65	98
height		13	JOE	13	M	63	105
weight		14	MICHAEL	13	M	58	95
▼ Rows		15	DAVID	13	M	59	79
All rows	40						

Launch an Analysis Platform

What is the distribution of the **weight** and **age** columns in the table?

- ☞ Click on the **Analyze** menu and choose the **Distribution** command.

This is called *launching* the Distribution platform. The launch dialog (**Figure 2.6**) now appears, prompting you to choose the variables you want to analyze.

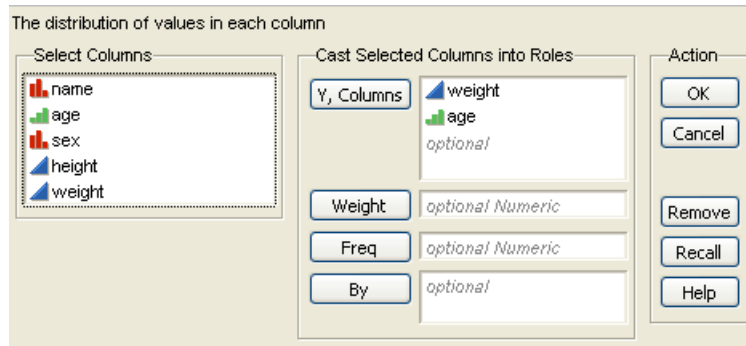
- ☞ Click on **weight** to highlight it in the variable list on the left of the dialog.
- ☞ Click **Y, Columns** to add it to the list of variables on the right of the dialog, which are the variables to be analyzed.
- ☞ Similarly, select the **age** variable and add it to the analysis variable list.

The term *variable* is often used to designate a column in the data table. Picking variables to fill roles is sometimes called *role assignment*.

You should now see the completed launch dialog shown in **Figure 2.6**.

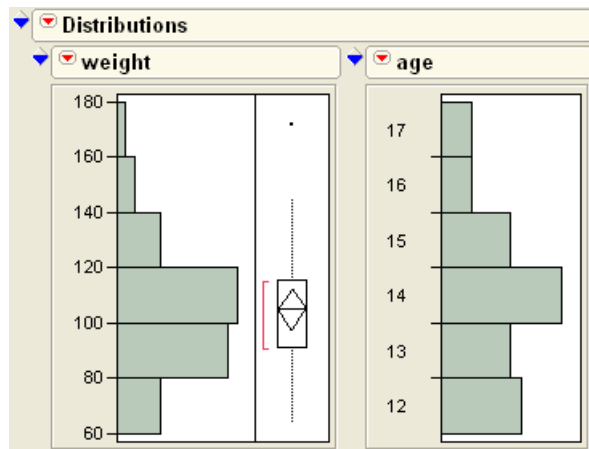
- ☞ Click **OK**, which closes the launch dialog and performs the Distribution analysis.

Figure 2.6 Distribution Launch Dialog



The resulting window shows the distribution of the two variables, weight and age, as in **Figure 2.7**.

Figure 2.7 Histograms from the Distribution Platform



Interact with the Surface of the Report

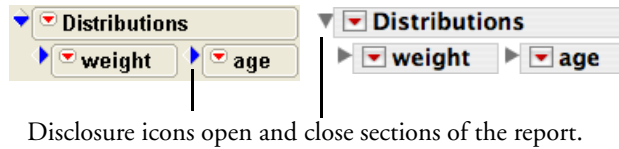
All JMP reports start with a basic analysis, which is then worked with interactively. This allows you to dig into a more detailed analysis, or customize the presentation. The report is a live object, not a dead transcript of calculations.

Row Highlighting

☞ Click on one of the histogram bars, for example, the **age** bar for 12-year-olds.

The bar is highlighted, along with portions of the bars in the other histogram and certain rows in the data table corresponding to the highlighted histogram bar. This is the dynamic

Figure 2.9 Disclosure Icons for Windows and Linux (left) and Macintosh (right)



Contextual Popup Menus

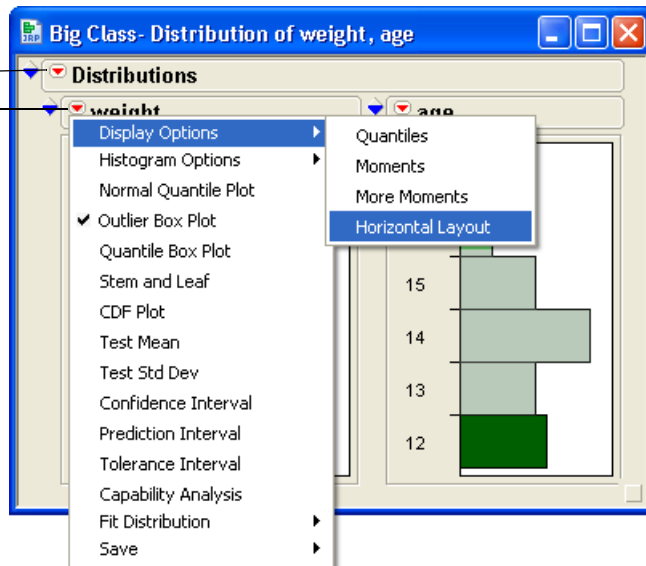
There is a small red triangle (a *hot spot*) on the title bar at the top of the analysis window that accesses popup menu commands for the analysis. This popup menu has commands specific to the platform. Hot spots on the title bars of each histogram contain commands that only influence that histogram. For example, you can change the orientation of the graphs in the Distribution platform by checking or unchecking **Display Options > Horizontal Layout** (Figure 2.10).

- ☞ Click on one of the menus next to weight or age and select **Display Options > Horizontal Layout**.

Figure 2.10 Display Options Menu

Click here for options relating to all histograms.

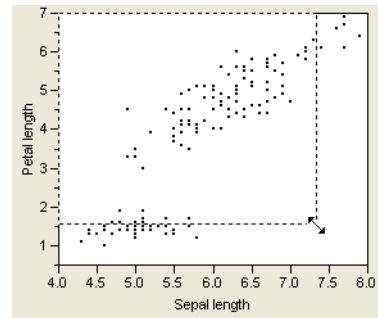
Click here to reveal a menu for each histogram.



In this same popup menu, you find options for performing further analyses or saving parts of the analysis in several forms. Whenever you see a red triangle hot spot, there are more options available. The options are specific to the context of the outline level where they are located. Many options are explained in later sections of this book.


Resizing Graphs


If you want to resize the graph windows in an analysis, move your mouse over the side or corner of the graph. The cursor changes to a double arrow, which lets you to drag the borders of the graph to the position you want.





Special Tools


When you need to do something special, pick a tool in the tools menu or tool palette and click or drag inside the analysis.


The grabber () is for grabbing objects.



 Select the grabber, then click and drag in a continuous histogram.



The brush () is for highlighting all the data in a rectangular area.


 Try getting the brush and dragging in the histogram. To change the size of the rectangle, option-drag (Macintosh), Alt-drag (Windows) or Shift-Alt-drag on Linux.


The lasso () is for selecting points by roping them in. We use this later in scatterplots.


The crosshairs () are for sighting along lines in a graph.

The magnifier () is for zooming in to certain areas in a plot. Hold down the  (Macintosh) Alt (Windows) or Shift+Alt (Linux) key and click to restore the original scaling.

The drawing tools () let you draw circles, squares, lines and shapes to annotate your report. The annotate tool () is for adding text annotations anywhere on the report.

The question mark () is for getting help on the analysis platform surface.

 Get the question mark tool and click on different areas in the Distribution platform.

The selection tool () is for picking out an area to copy so that you can paste its contents into another application. Hold down the Shift key to select multiple report sections. Refer to the chapter “Data Tables, Reports, and Scripts” on page 27 for details.

In JMP, the surface of an analysis platform bristles with interactivity. Launching an analysis is just the starting point. You then explore, evaluate, follow clues, dig deeper, get more details, and fine-tune the presentation.

Modeling Type

Notice in the previous example that there are different kinds of graphs and reports for **weight** and **age**. This is because the variables are assigned different *modeling types*. The **weight** column has a *continuous* modeling type, so JMP treats the numbers as values from a continuous scale. The **age** column has an *ordinal* modeling type, so JMP treats its values as labels of discrete categories.

Here is a brief description of the three modeling types:

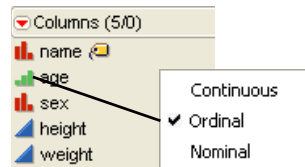
- *Continuous* (📊) are numeric values used directly in an analysis.
- *Ordinal* (📈) values are category labels, but their order is meaningful.
- *Nominal* (📊) values are treated as unordered, categorical names of levels.

The ordinal and nominal modeling types are treated the same in most analyses, and are often referred to collectively as *categorical*.

You can change the modeling type using the Columns panel at the left of the data grid (**Figure 2.11**). Notice the 📈 beside the column heading for **age**. This icon is a popup menu.

🖱️ Click on the 📈 to see the menu for choosing the modeling type for a column.

Figure 2.11 Modeling Type Popup Menu on the Columns Panel



Why does JMP distinguish among modeling types? For one thing, it's a convenience feature. You are telling JMP ahead of time how you want the column treated so that you don't have to say it again every time you do an analysis. It also helps reduce the number of commands you need to learn. Instead of two distribution platforms, one for continuous variables and a different one for categorical variables, a single command performs the anticipated analysis based on the modeling type you assigned.

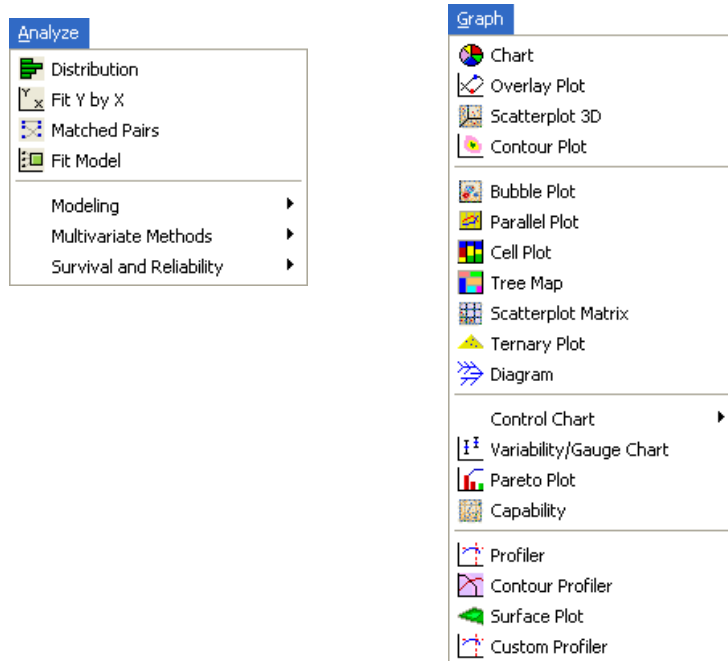
You can change the modeling type whenever you want the variable treated differently. For example, if you wanted to find the mean of **age** instead of categorical frequency counts, simply change the modeling type from ordinal to continuous and repeat the analysis.

The following sections demonstrate how the modeling type affects the kind of analysis from several platforms.

Analyze and Graph

The **Analyze** and **Graph** menus, shown here, launch interactive platforms to analyze data.

Figure 2.12 Analyze and Graph Menus



The **Analyze** menu is for statistics and data analysis. The **Graph** menu is for specialized plots. That distinction, however, doesn't prevent analysis platforms from being full of graphs, nor the graph platforms from computing statistics. Each platform provides a context for sets of related statistical methods and graphs. It won't take long to learn this short list of platforms. The next sections briefly describe the **Analyze** and **Graph** commands.

The Analyze Menu

Distribution is for univariate statistics, and describes the distribution of values for each variable, one at a time, using histograms, box plots, and other statistics.

Fit Y by X is for bivariate analysis. A bivariate analysis describes the distribution of a y -variable as it depends on the value of the x -variable. The continuous or categorical modeling type of

the y - and x - variables leads to one of the four following analyses: scatterplot with regression curve fitting, one-way analysis of variance, contingency table analysis, or logistic regression.

Matched Pairs compares means between two response columns using a paired t -test. Often the two columns represent measurements on the same subject before and after some treatment.

Fit Model launches a general fitting platform for linear models. Analyses found in this platform include multiple regression, analysis of variance models, generalized linear models, and logistic regression.

Modeling

Screening helps select a model to fit to a two-level screening design by showing which effects are large.

Nonlinear fits models that are nonlinear in their parameters, using iterative methods.

Neural Net implements a standard type of neural network.

Gaussian Process models the relationship between a continuous response and one or more continuous predictors. These models are common in areas like computer simulation experiments, such as the output of finite element codes, and they often perfectly interpolate the data. Gaussian processes can deal with these no-error-term models.

Time Series lets you explore, analyze, and forecast univariate time series taken over equally spaced time periods. The analysis begins with a plot of the points in the time series with autocorrelations and partial autocorrelations, and can fit ARIMA, seasonal ARIMA, transfer function models, and smoothing models.

Partition recursively partitions values, similar to CART™ and CHAID™.

Categorical tabulates and summarizes categorical response data, including multiple response data, and calculates test statistics. It is designed to handle survey and other categorical response data, including multiple response data like defect records, side effects, and so on.

Multivariate Methods

Multivariate describes relationships among variables, focusing on the correlation structure: correlations and other measures of association, scatterplot matrices, multivariate outliers, and principal components.

Cluster allows for k -means and hierarchical clustering. Normal mixtures and Self-Organizing Maps (SOMs) are found in this platform.

Principal Components derives a small number of independent linear combinations (principal components) of a set of variables that capture as much of the variability in the original variables as possible. JMP offers several types of orthogonal and oblique Factor-Analysis-Style rotations to help interpret the extracted components.

Discriminant fits discriminant analysis models, categorizing data into groups.

PLS implements partial least-squares analyses.

Item Analysis analyzes questionnaire or test data using Item Response Theory.

Survival and Reliability

Survival /Reliability models the time until an event, allowing censored data. This kind of analysis is used in both reliability engineering and survival analysis.

Fit Parametric Survival opens the Fit Model dialog to model parametric (regression) survival curves.

Fit Proportional Hazards opens the Fit Model dialog to fit the Cox proportional hazards model.

Recurrence Analysis analyzes repairable systems.

The Graph Menu

Chart gives many forms of charts such as bar, pie, line, and needle charts.

Overlay Plot overlays several numeric y -variables, with options to connect points, or show a step plot, needle plot, or others. It is possible to have two y -axes in these plots.

Scatterplot 3D produces a three-dimensional spinnable display of values from any three numeric columns in the active data table. It also produces an approximation to higher dimensions through principal components, standardized principal components, rotated components, and biplots.

Contour Plot constructs a contour plot for one or more response variables for the values of two x -variables. **Contour Plot** assumes the x values lie in a rectangular coordinate system, but the observed points do not have to form a grid.

Bubble Plot draws a scatter plot which represents its points as circles (bubbles). Optionally the bubbles can be sized according to another column, colored by yet another column, aggregated across groups defined by one or more other columns, and dynamically indexed by a time column.

Parallel Plot shows connected-line plots of several variables at once.

Cell Plot produces a “heat map” of a column, assigning colors based on a gradient (for continuous variables) or according to a discrete list of colors (for categorical variables).

Tree Map presents a two-dimensional, tiled view of the data.

Scatterplot Matrix produces scatterplot matrices.

Ternary Plot constructs a plot using triangular coordinates. The ternary platform uses the same options as the contour platform for building and filling contours. In addition, it uses a specialized crosshair tool that lets you read the triangular axis values.

Diagram is used to construct *Ishikawa charts*, also called *fishbone charts*, or *cause-and-effect diagrams*. These charts are useful when organizing the sources (causes) of a problem (effect), perhaps for brainstorming, or as a preliminary analysis to identify variables in preparation for further experimentation.

Control Chart presents a submenu of various control charts available in JMP.

Variability/Gage Chart is used for analyzing measurement systems. Data can be continuous measurements or attributes.

Pareto Plot creates a bar chart (Pareto chart) that displays the severity (frequency) of problems in a quality-related process or operation. Pareto plots compare quality-related measures or counts in a process or operation. The defining characteristic of Pareto plots is that the bars are in descending order of values, which visually emphasizes the most important measures or frequencies.

Capability measures the conformance of a process to given specification limits. Using these limits, you can compare a current process to specific tolerances and maintain consistency in production. Graphical tools such as the goalpost plot and box plot give you quick visual ways of observing within-spec behaviors.

Profiler is available for tables with columns whose values are computed from model prediction formulas. Usually, profiler plots appear in standard least squares reports, where they are a menu option. However, if you save the prediction equation from the analysis, you can access the prediction profile independent of a report from the **Graph** menu and look at the model using the response column with the saved prediction formula.

Contour Profiler works the same as the **Profiler** command. It is usually accessed from the Fit Model platform when a model has multiple responses. However, if you save the prediction

formulas for the responses, you can access the Contour Profiler at a later time from the **Graph** menu and specify the columns with the prediction equations as the response columns.

Surface Plot draws up to four three-dimensional, rotatable surfaces. It can also produce density shells and isosurfaces.

Custom Profiler is an advanced feature for optimization and simulation.

Navigating Platforms and Building Context

The first few times JMP is used, most people have navigational questions: How do I get a particular graph? How do I produce a histogram? How do I get a t -test?

The strategy for approaching JMP analyses is to build an analysis context. Once you build that context, the graphs and statistics become easily available—often they happen automatically, without having to ask for them specifically.

There are three keys for establishing the context:

- Designating the *Modeling Type* of the variables in the analysis as either continuous, ordinal, or nominal.
- Assigning *X or Y Roles* to identify whether the variable is a response (Y) or a factor (X).
- Selecting an *analysis platform* for the general approach and character of the analysis.

Once you settle on a context, commands appear in logical places.

Contexts for a Histogram

Suppose you want to display a histogram. In other software, you might find a histogram command in a graph menu, but in JMP you need to think of the context. You want a histogram so that you can see a distribution of values. So, launch the **Distribution** platform in the **Analyze** menu. Once launched, there are many graphs and reports available for focusing on the distribution of values.

Occasionally, you may want the histogram as a presentation graph. Then, instead of using the Distribution platform, use the Chart platform in the **Graph** menu.

Contexts for the t -Test

Suppose you want a t -test. Other software might have a t -test command on a main menu. JMP has many t -test commands, because there are many contexts in which this statistic is used. So first, you have to build the context of your situation.

If you want the t -test to test a single variable's mean against a hypothesized value, you are focusing on a univariate distribution, and therefore launch the Distribution platform (**Analyze > Distribution**). On the title bar of the distribution report is a popup menu with the command **Test Mean**. This command gives you a t -test, as well as the option to conduct a nonparametric test.

If you want the t -test to compare the means of two independent groups, then you have two variables in the context—perhaps a continuous Y response and a categorical X factor. Since the analysis deals with two variables, use the Fit Y By X platform. If you launch the Fit Y by X platform, you'll see the side-by-side comparison of the two distributions, and you can use the **t test** or **Means/Anova/Pooled t** command from the popup menu on the analysis title bar.

If you want to compare the means of two continuous responses that form matched pairs, there are several ways to build the appropriate context. You can make a third data column to form the difference of the responses, and use the Distribution platform to do a t -test that the mean of the differences is zero. Alternatively, you can use the **Matched Pairs** command to launch the Matched Pairs platform for the two variables. In Chapter 8, “The Difference between Two Means,” you will learn more ways to do a t -test.

Contexts for a Scatterplot

Suppose you want a scatterplot of two variables. The general context is a bivariate analysis, which suggests using the Fit Y by X platform. With two continuous variables, the Fit Y By X platform produces a scatterplot. You can then fit regression lines or other appropriate items with this scatterplot from the same report.

You might also consider the **Overlay Plot** command in the **Graph** menu when you want a presentation graph. As a **Graph** menu platform, it does not compute regressions, but can overlay multiple Y 's in the same graph and support two y -axes.

If you have a whole series of scatterplots for many variables in mind, your context is many bivariate associations, which is part of either the Multivariate or Scatterplot 3D platforms. A matrix of scatterplots appears automatically for Multivariate platform analyses.

Contexts for Nonparametric Statistics

There is not a separate platform for nonparametric statistics. However, there are many standard nonparametric statistics in JMP, positioned by context. When you test a mean in the Distribution platform, you have the option to do a (nonparametric) Wilcoxon signed-rank test. When you do a t -test or one-way ANOVA in the Fit Y by X platform, you also have three optional nonparametric tests, including the Wilcoxon rank sum (which is equivalent to the

Mann-Whitney U -test). If you want a nonparametric measure of association, like Kendall's τ or Spearman's correlation, look in the Multivariate platform.

The Personality of JMP

Here are some reasons why JMP is different from other statistical software:

Graphs are in the service of statistics (and vice versa). The goal of JMP is to provide a graph for every statistic, presented with the statistic. The graphs shouldn't appear in separate windows, but rather should work together. In the analysis platforms, the graphs tend to follow the statistical context. In the graph platforms, the statistics tend to follow the graphical context.

JMP encourages good data analysis. In the example presented in this chapter, you didn't have to ask for a histogram because it appeared when you launched the Distribution platform. The Distribution platform was designed that way, because in good data analysis you always examine a graph of a distribution before you start doing statistical tests on it. This encourages responsible data analysis.

JMP allows you to make discoveries. JMP was developed with the charter to be "Statistical Discovery Software." After all, you want to find out what you didn't know, as well as try to prove what you already know. Graphs attract your attention to an outlier or other unusual feature of the data that might prove valuable to discovery. Imagine Marie Curie using a computer for her pitchblende experiment. If software had given her only the end results, rather than showing her the data and the graphs, she might not have noticed the discrepancy that led to the discovery of radium.

JMP bristles with interactivity. In some products, you have to specify exactly what you want ahead of time because often that is your only chance while doing the analysis. JMP is interactive, so everything is open to change and customization at any point in the analysis. It is easier to remove a histogram when you don't want it than decide ahead of time that you want one.

You can see your data from multiple perspectives. Did you know that a t -test for two groups is a special case of an F -test for several groups? With JMP, you tend to get general methods that are good for many situations, rather than specialty methods for special cases. You also tend to get several ways to test the same thing. For two groups, there is a t -test and its equivalent F -test. When you are ready for more, there are three nonparametric tests to use in the same situation. You also can test for different variances across the groups and get appropriate results. And there are two graphs to show you the separation of the means. Even after you perform statistical tests, there are multiple ways of looking at the results, in terms of the p -

value, the confidence intervals, least significant differences, the sample size, and least significant number. With this much statistical breadth, it is good that commands appear as you qualify the context, rather than having to select multiple commands from a single menu bar. JMP unfolds the details progressively, as they become relevant.

