Release 8

Discovering JMP

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

8.0.2

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2009. *JMP® 8 Discovering JMP*. Cary, NC: SAS Institute Inc.

JMP® 8 Discovering JMP

ISBN 978-1-60764-318-0

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, December 2009

Contents

Discovering JMP

Credits and Acknowledgments

Origin

JMP was developed by SAS Institute Inc., Cary, NC. JMP is not a part of the SAS System, though portions of JMP were adapted from routines in the SAS System, particularly for linear algebra and probability calculations. Version 1 of JMP went into production in October, 1989.

Credits

JMP was conceived and started by John Sall. Design and development were done by John Sall, Chung-Wei Ng, Michael Hecht, Richard Potter, Brian Corcoran, Annie Dudley Zangi, Bradley Jones, Craige Hales, Chris Gotwalt, Paul Nelson, Xan Gregg, Jianfeng Ding, Eric Hill, John Schroedl, Laura Lancaster, Scott McQuiggan, and Peng Liu.

In the SAS Institute Technical Support division, Wendy Murphrey and Toby Trott provide technical support and conduct test site administration. Statistical technical support is provided by Craig DeVault, Duane Hayes, Elizabeth Edwards, Kathleen Kiernan, and Tonya Mauldin.

Nicole Jones, Jim Borek, Kyoko Keener, Hui Di, Joseph Morgan, Wenjun Bao, Fang Chen, Susan Shao, Hugh Crews, Yusuke Ono and Kelci Miclaus provide ongoing quality assurance. Additional testing and technical support is done by Noriki Inoue, Kyoko Takenaka, and Masakazu Okada from SAS Japan.

Bob Hickey is the release engineer.

The JMP manuals were written by Ann Lehman, Lee Creighton, John Sall, Bradley Jones, Erin Vang, Melanie Drake, Meredith Blackwelder, Diane Perhac, Jonathan Gatlin, and Susan Conaghan with contributions from Annie Dudley Zangi and Brian Corcoran. Creative services and production was done by SAS Publications. Melanie Drake implemented the help system.

Jon Weisz and Jeff Perkinson provided project management. Also thanks to Lou Valente, Ian Cox, Mark Bailey, and Malcolm Moore for technical advice.

Thanks also to Georges Guirguis, Warren Sarle, Gordon Johnston, Duane Hayes, Russell Wolfinger, Randall Tobias, Robert N. Rodriguez, Ying So, Warren Kuhfeld, George MacKensie, Bob Lucas, Warren Kuhfeld, Mike Leonard, and Padraic Neville for statistical R&D support. Thanks are also due to Doug Melzer, Bryan Wolfe, Vincent DelGobbo, Biff Beers, Russell Gonsalves, Mitchel Soltys, Dave Mackie, and Stephanie Smith, who helped us get started with SAS Foundation Services from JMP.

Acknowledgments

We owe special gratitude to the people that encouraged us to start JMP, to the alpha and beta testers of JMP, and to the reviewers of the documentation. In particular we thank Michael Benson, Howard Yetter (d), Andy Mauromoustakos, Al Best, Stan Young, Robert Muenchen, Lenore Herzenberg, Ramon Leon, Tom Lange, Homer Hegedus, Skip Weed, Michael Emptage, Pat Spagan, Paul Wenz, Mike Bowen, Lori Gates, Georgia Morgan, David Tanaka, Zoe Jewell, Sky Alibhai, David Coleman,

Linda Blazek, Michael Friendly, Joe Hockman, Frank Shen, J.H. Goodman, David Iklé, Barry Hembree, Dan Obermiller, Jeff Sweeney, Lynn Vanatta, and Kris Ghosh.

Also, we thank Dick DeVeaux, Gray McQuarrie, Robert Stine, George Fraction, Avigdor Cahaner, José Ramirez, Gudmunder Axelsson, Al Fulmer, Cary Tuckfield, Ron Thisted, Nancy McDermott, Veronica Czitrom, Tom Johnson, Cy Wegman, Paul Dwyer, DaRon Huffaker, Kevin Norwood, Mike Thompson, Jack Reese, Francois Mainville, and John Wass.

We also thank the following individuals for expert advice in their statistical specialties: R. Hocking and P. Spector for advice on effective hypotheses; Robert Mee for screening design generators; Roselinde Kessels for advice on choice experiments; Greg Piepel, Peter Goos, J. Stuart Hunter, Dennis Lin, Doug Montgomery, and Chris Nachtsheim for advice on design of experiments; Jason Hsu for advice on multiple comparisons methods (not all of which we were able to incorporate in JMP); Ralph O'Brien for advice on homogeneity of variance tests; Ralph O'Brien and S. Paul Wright for advice on statistical power; Keith Muller for advice in multivariate methods, Harry Martz, Wayne Nelson, Ramon Leon, Dave Trindade, Paul Tobias, and William Q. Meeker for advice on reliability plots; Lijian Yang and J.S. Marron for bivariate smoothing design; George Milliken and Yurii Bulavski for development of mixed models; Will Potts and Cathy Maahs-Fladung for data mining; Clay Thompson for advice on contour plotting algorithms; and Tom Little, Damon Stoddard, Blanton Godfrey, Tim Clapp, and Joe Ficalora for advice in the area of Six Sigma; and Josef Schmee and Alan Bowman for advice on simulation and tolerance design.

For sample data, thanks to Patrice Strahle for Pareto examples, the Texas air control board for the pollution data, and David Coleman for the pollen (eureka) data.

Translations

Erin Vang, Trish O'Grady, Elly Sato, and Kyoko Keener coordinate localization. Special thanks to Noriki Inoue, Kyoko Takenaka, Masakazu Okada, Naohiro Masukawa and Yusuke Ono (SAS Japan); and Professor Toshiro Haga (retired, Tokyo University of Science) and Professor Hirohiko Asano (Tokyo Metropolitan University) for reviewing our Japanese translation; Professors Fengshan Bai, Xuan Lu, and Jianguo Li at Tsinghua University in Beijing, and their assistants Rui Guo, Shan Jiang, Zhicheng Wan, and Qiang Zhao; and William Zhou (SAS China) and Zhongguo Zheng, professor at Peking University, for reviewing the Simplified Chinese translation; Jacques Goupy (consultant, ReConFor) and Olivier Nuñez (professor, Universidad Carlos III de Madrid) for reviewing the French translation; Dr. Byung Chun Kim (professor, Korea Advanced Institute of Science and Technology) and Duk-Hyun Ko (SAS Korea) for reviewing the Korean translation; Bertram Schäfer and David Meintrup (consultants, StatCon) for reviewing the German translation; Patrizia Omodei, Maria Scaccabarozzi, and Letizia Bazzani (SAS Italy) for reviewing the Italian translation. Finally, thanks to all the members of our outstanding translation teams.

Past Support

Many people were important in the evolution of JMP. Special thanks to David DeLong, Mary Cole, Kristin Nauta, Aaron Walker, Ike Walker, Eric Gjertsen, Dave Tilley, Ruth Lee, Annette Sanders, Tim Christensen, Jeff Polzin, Eric Wasserman, Charles Soper, Wenjie Bao, and Junji Kishimoto. Thanks to SAS Institute quality assurance by Jeanne Martin, Fouad Younan, and Frank Lassiter. Additional testing for Versions 3 and 4 was done by Li Yang, Brenda Sun, Katrina Hauser, and Andrea Ritter.

Also thanks to Jenny Kendall, John Hansen, Eddie Routten, David Schlotzhauer, and James Mulherin. Thanks to Steve Shack, Greg Weier, and Maura Stokes for testing JMP Version 1.

Thanks for support from Charles Shipp, Harold Gugel (d), Jim Winters, Matthew Lay, Tim Rey, Rubin Gabriel, Brian Ruff, William Lisowski, David Morganstein, Tom Esposito, Susan West, Chris Fehily, Dan Chilko, Jim Shook, Ken Bodner, Rick Blahunka, Dana C. Aultman, and William Fehlner.

Technology License Notices

The ImageMan DLL is used with permission of Data Techniques, Inc.

About This Guide

Discovering JMP provides a general introduction to the JMP software. This guide assumes that you have no knowledge of JMP. Whether you are an analyst, researcher, student, professor, or statistician, this guide gives you a general overview of JMP's user interface and features.

This guide introduces you to the following information:

- Starting JMP
- The structure of a JMP window
- Getting data into JMP
- Preparing and manipulating data
- Using interactive graphs to learn from your data
- Performing simple analyses to augment graphs
- Customizing JMP and special features

This guide contains six chapters. Each chapter contains examples that reinforce the concepts presented in the chapter. All of the statistical concepts are at an introductory level. The sample data used in this book are included with the software. Here is a description of each chapter:

- Chapter 1, Introducing JMP, provides an overview of the JMP application. You learn how content is organized and how to navigate the software.
- Chapter 2, Preparing Data, describes how to import data from a variety of sources, and prepare it for analysis. There is also an overview of data manipulation tools.
- Chapter 3, Visualizing Your Data, describes graphs and charts you can use to visualize and understand your data. The examples range from simple analyses involving a single variable, to multi-variable graphs that enable you to see relationships between many variables.
- Chapter 4, Analyzing Your Data, describes many commonly used analysis techniques. These range from simple techniques that do not require the use of statistical methods, to advanced techniques, where knowledge of statistics is useful.
- Chapter 5, Customizing JMP, describes using journals and projects, settings preferences, and includes a section on the JMP Scripting Language (JSL).
- Chapter 6, Special Features, describes how to automatically update graphs and analyses as data changes, and how JMP interacts with SAS.

After reviewing this guide, you will be comfortable navigating and working with your data in JMP.

While JMP is available in both Windows and Mac operating systems, the material in this guide is based on a Windows operating system.

Chapter 1

Introducing JMP

Basic Concepts

Welcome to JMP, pronounced “jump.”

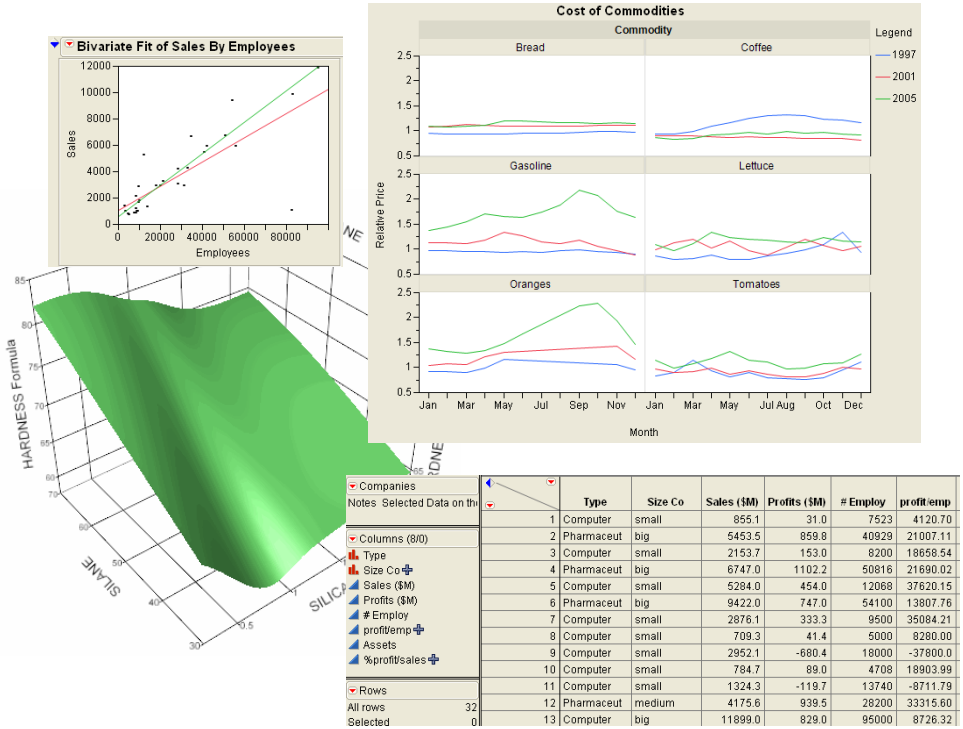
JMP software is a powerful, interactive, data visualization and statistical analysis tool. You can use the understanding you gain from the visualization tools to guide your analytics and statistical analyses.

JMP is equally useful to the researcher who wants to perform a wide range of statistical analyses and modeling, and the business analyst who wants to quickly uncover trends and patterns in data. With JMP, you do not have to be an expert in statistics to get information from your data.

For example, you can use JMP to:

- Create interactive graphs and charts to explore your data and discover relationships
- Discover patterns of variation across many variables at once
- Explore and summarize large amounts of data
- Develop powerful statistical models you can use to predict the future

Figure 1.1 Examples of JMP Elements



	Type	Size Co	Sales ($M)	Profits ($M)	# Employ	profit/emp
1	Computer	small	855.1	31.0	7523	4120.70
2	Pharmaceut	big	5453.5	859.8	40929	21007.11
3	Computer	small	2153.7	153.0	8200	18658.54
4	Pharmaceut	big	6747.0	1102.2	50816	21690.02
5	Computer	small	5284.0	454.0	12068	37620.15
6	Pharmaceut	big	9422.0	747.0	54100	13807.76
7	Computer	small	2876.1	333.3	9500	35084.21
8	Computer	small	709.3	41.4	5000	8280.00
9	Computer	small	2952.1	-680.4	18000	-37800.0
10	Computer	small	784.7	89.0	4708	18903.99
11	Computer	small	1324.3	-119.7	13740	-8711.79
12	Pharmaceut	medium	4175.6	939.5	28200	33315.60
13	Computer	big	11899.0	829.0	95000	8726.32

Contents

How Do I Get Started?

This section discusses how to start a JMP session and familiarizes you with the user interface.

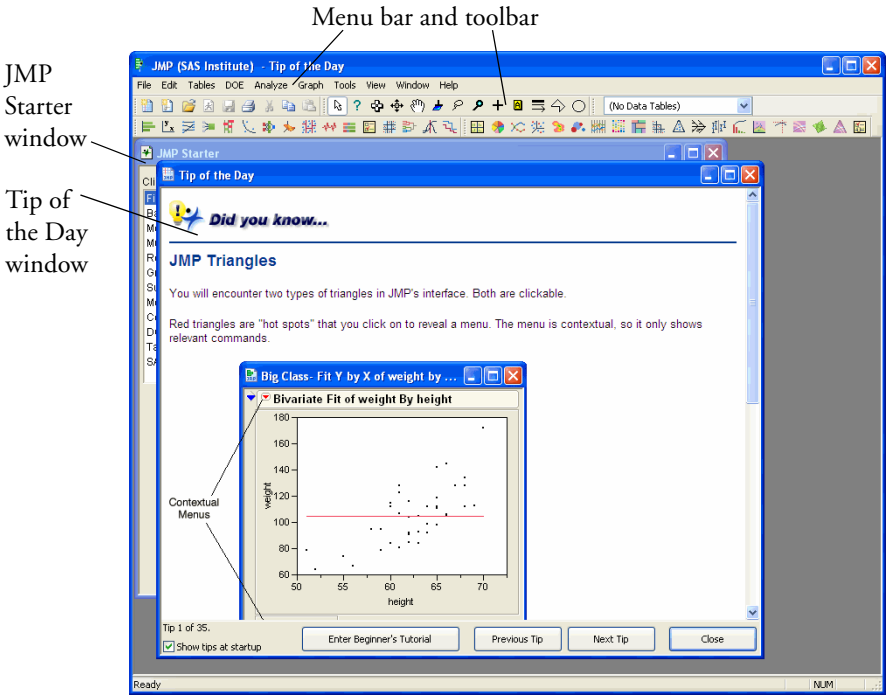
Starting JMP

Here are two common ways to start JMP:

- Double-click the JMP icon, usually found on your desktop. This starts JMP, but does not open any existing JMP files.
- Double-click an existing JMP file. This starts JMP and opens the file.

If you do not open an existing JMP file, your initial view is shown in Figure 1.2.

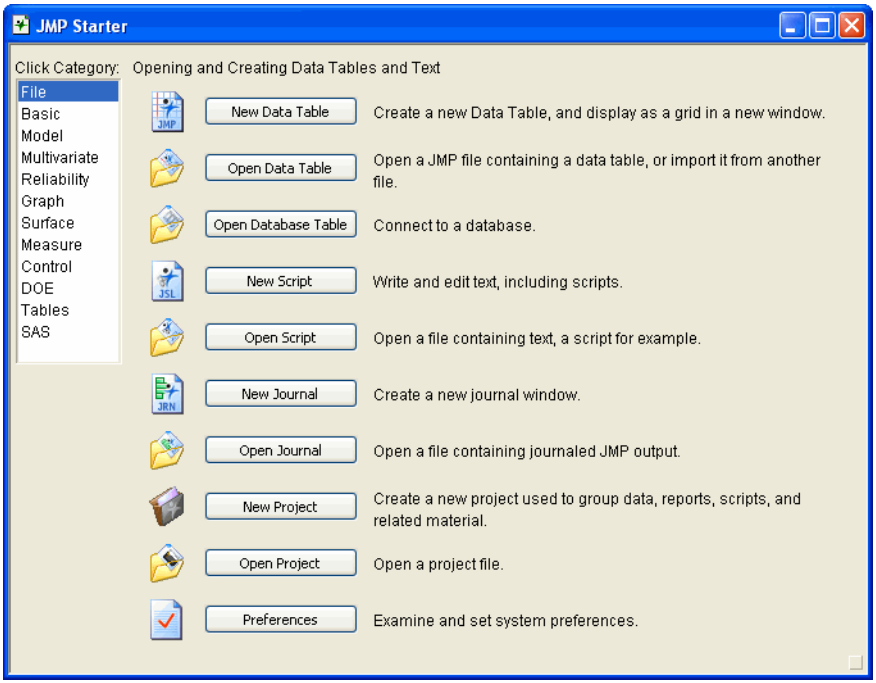
Figure 1.2 Initial View of JMP



The initial view of JMP includes the Tip of the Day window and the JMP Starter window. At the top of the main JMP window is a menu bar and a toolbar. You can access most JMP features through both the menus and toolbars. By default, windows in JMP are not maximized. This enables you to see the interaction between the windows.

For the initial view, the active window (top-most window) is the Tip of the Day window. Close that window by clicking the close button at the top right of the window. The JMP Starter is now the active window, and is shown in Figure 1.3.

Figure 1.3 JMP Starter



The JMP Starter is another way to access most of JMP's platforms and features. On the left is a list of categories. On the right are the features and commands related to that category. See the Appendix for a description of most of these features.

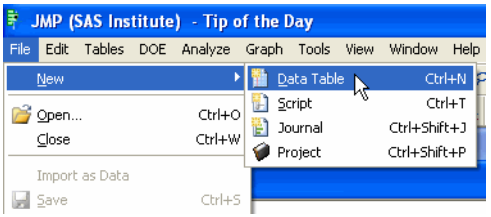
Creating a New Data Table

To create a new data table, perform one of the following actions (see Figure 1.4).

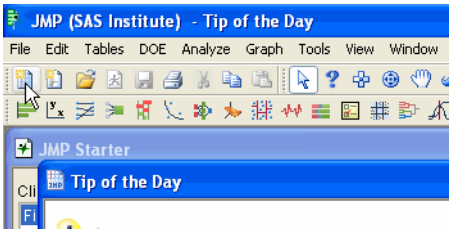
- Select **File > New** > **Data Table**.
- On the toolbar, click the **Data Table** icon.
- In the JMP Starter window, click the **File** category on the left, and then click the **New Data Table** button on the right.

Figure 1.4 Three Ways to Create a New Data Table

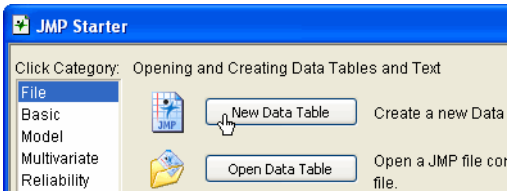
Create a new data table using the menu bar



Create a new data table using the toolbar



Create a new data table using the JMP Starter



Notice that the icon on the menu item matches the icon on the toolbar item. This is true for all menu and toolbar items that perform corresponding actions.

Once you have created a data table, you can add your data. See "Moving Data into JMP," p. 19.

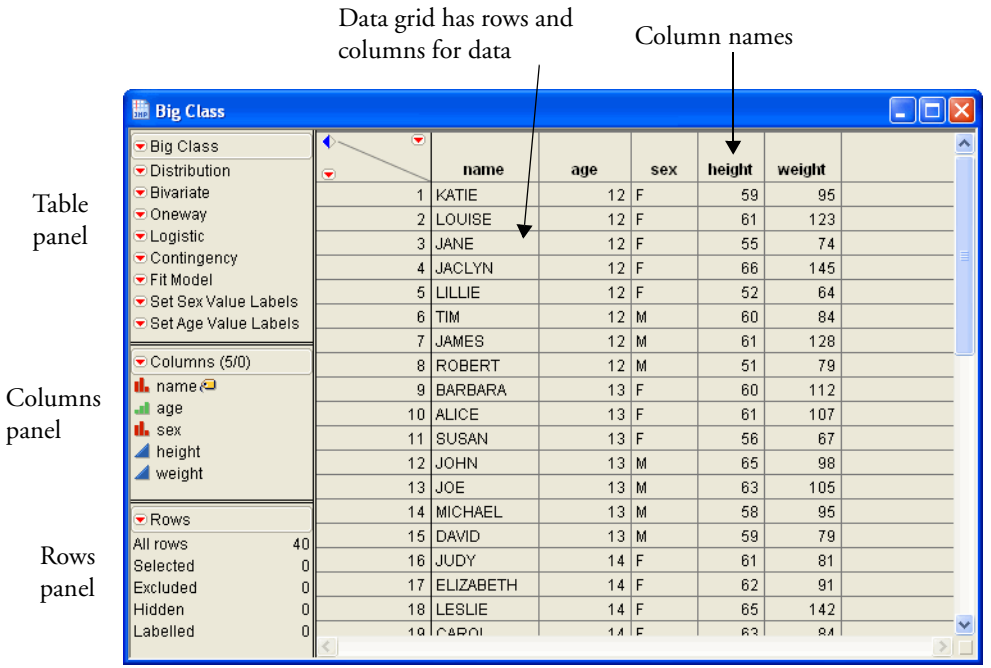
Understanding JMP Data Tables

A JMP data table is a collection of data organized into rows and columns. It is similar to an Excel spreadsheet, but with some important differences that are discussed later. A JMP data table also might also contain other information like notes, variables, and scripts. These supplementary items are discussed in later chapters.

You can open an example JMP data table and see the parts. Follow the steps below to open the Big Class.jmp data table.

1 From the **Help** menu, select **Sample Data**.
2 In the **Teaching Demonstrations** category, open the **Examples for Teaching** node by selecting the blue triangle next to it.
3 Click the **Big Class** link to open the Big Class.jmp data table. A partial view of the data table is shown in Figure 1.5.

Figure 1.5 Big Class.jmp Data Table



The major parts of a JMP data table are:

Data grid The data grid, or spreadsheet, contains the actual data. It is divided by rows and columns. Generally, each row of a spreadsheet is a data point, and the different columns (or variables) give information about the data point. In Figure 1.5, each row corresponds to a person, and there are five columns of information. The information given about the person includes name, age, sex, height, and weight. Each column has a header, or name. That name is not part of the spreadsheet's total count of rows. In this example, the column names are name, age, sex, height, and weight.

Table panel The table panel can contain table variables or table scripts. In this example, there are several saved scripts that are used to automatically recreate analyses. Table variables and table scripts are discussed in a later chapter.

Columns panel The columns panel indicates the number of columns and lists the columns. In this example there are five columns. Icons to the left of the column name indicate a column's modeling type. Icons to the right indicate any attributes assigned to the column. These icons are discussed in a later chapter.

Rows panel The rows panel indicates the number of rows in the data table, and how many rows are selected, excluded, hidden, or labeled. In this example there are 40 rows in the data table.

Interacting with the data grid, which includes adding rows and columns, entering data, and editing data, is discussed in the next chapter. When more than one JMP data table is viewed at the same time, each one appears in a separate window.

Using Platforms

Once data is in a data table, you can create any of JMP's graphs or plots, and do many analyses. All of JMP's features are accessed through platforms, which are found primarily on the **Analyze** or **Graph** menus. They are called platforms because they do not just produce simple static output. Platform results appear in separate report windows, are highly interactive, and are linked to the data table.

The platforms under the **Analyze** and **Graph** menus provide a variety of analytical features and data exploration tools (graphs and plots). Each platform is described in the Appendix.

Launching a Platform and Viewing Results

To see an example of using a JMP platform, you can use the Companies.jmp file. If you want to do a basic analysis of the variable Profits ($M), use the **Distribution** platform under the **Analyze** menu. To do the analysis, follow the steps below:

1 Open the launch window by selecting **Analyze** > **Distribution**.

2 Assign the Profits ($M) variable to the **Y, Columns** role. To do this, select Profits ($M) from the Select Columns box and click the **Y, Columns** button. The variable Profits ($M) appears in the **Y, Columns** role. See Figure 1.6 for the completed window. Alternatively, you can click and drag Profits ($M) from the Select Columns box to the **Y, Columns** box.

3 Click **OK**. The Distribution report window appears. See Figure 1.7.

Figure 1.6 Assign Profits ($M)

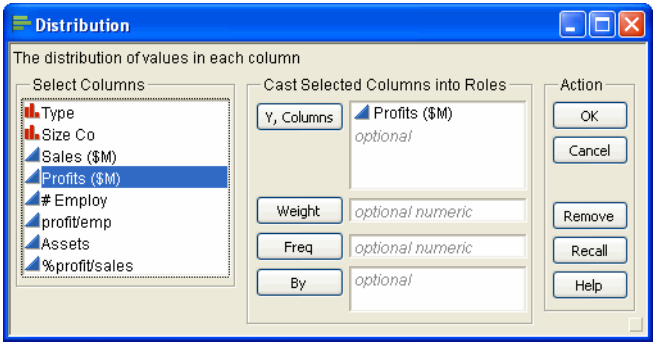
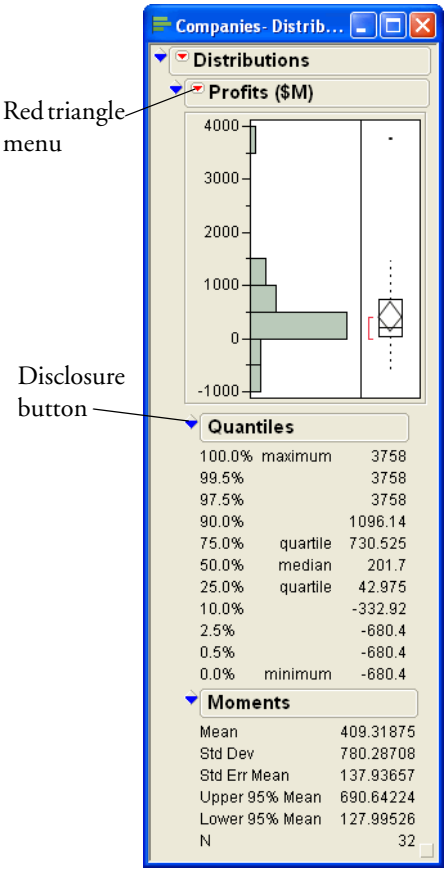


Figure 1.7 Initial Distribution Report Window



The report window contains basic plots or graphs and a preliminary analysis. The results appear in an outline format, and you can show or hide any section by clicking on the disclosure button. Red triangle menus contain options and commands that let you request additional graphs and analysis at any time.

Adding and Removing Analyses and Graphics

The report window initially contains basic graphs and statistics. Use the red triangle menus to:

- Remove existing graphs or statistics
- Request additional graphs or statistics

As an example of removing existing content, follow the steps below to remove the outlier box plot:

1 Select the red triangle menu next to Profits ($M). See Figure 1.8.
2 Uncheck the **Outlier Box Plot** option. The outlier box plot is removed from the report window. See Figure 1.9.

Figure 1.8 Red Triangle Menu Next to Profits ($M)

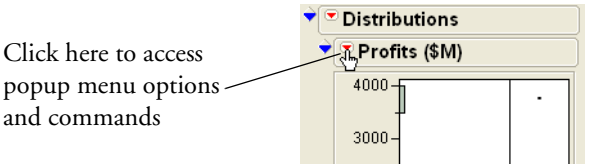
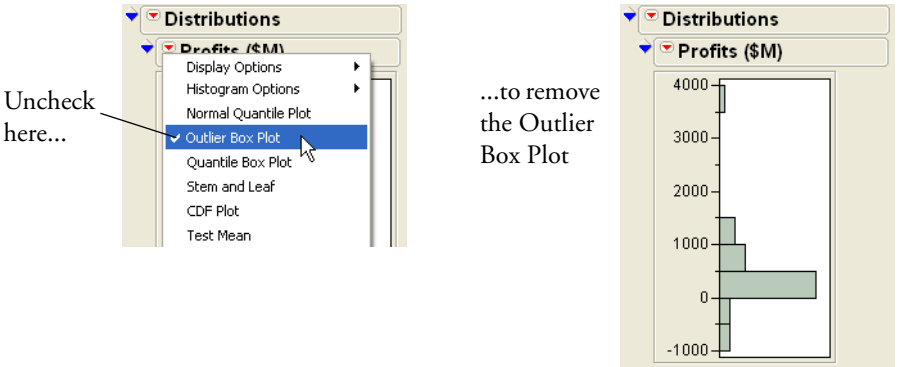


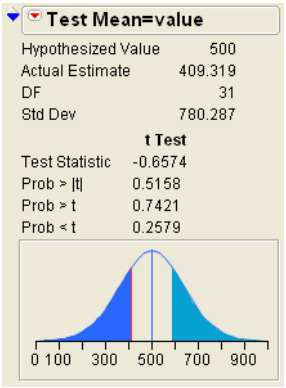
Figure 1.9 Removing the Outlier Box Plot



As an example of requesting additional output, follow the steps below to add a hypothesis test:

1 Select the **Test Mean** command on the same red triangle menu.
2 Fill out the resulting window by typing 500 in the **Specify Hypothesized Mean** box.
3 Click **OK**. The test for the mean is added to the report. See Figure 1.10.

Figure 1.10 Test for the Mean



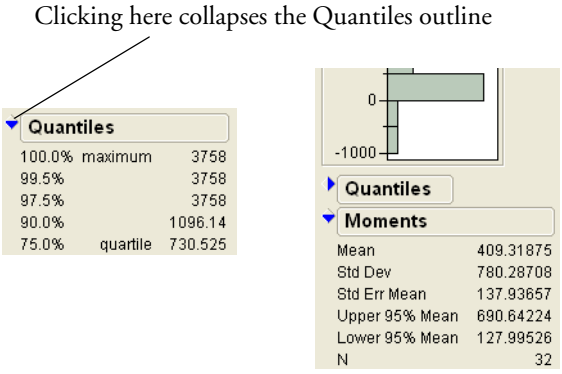
Interacting with Platform Results

All platforms produce results that are interactive in many ways. Two of the common ways you can interact with platform results include the following:

- Closing or opening outlines
- Connecting platform results to the data table

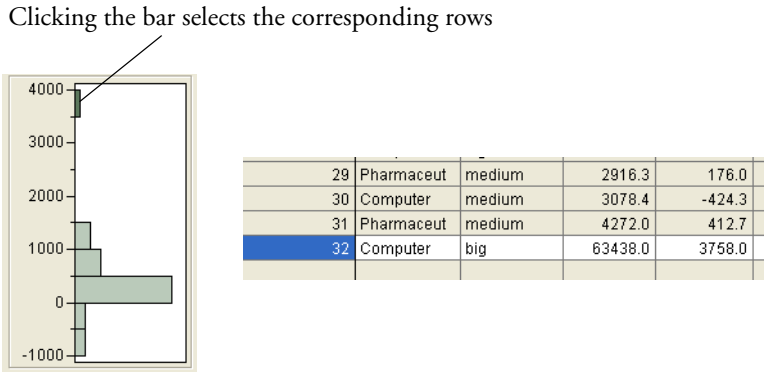
You can close or open sections (or outlines) of the platform results. For example, to close the **Quantiles** outline, click the disclosure button next to **Quantiles**. See Figure 1.11.

Figure 1.11 Close the Quantiles Outline



Platform results are interconnected with the data table. The histogram shows that there are a few students that are shorter than the rest. You can quickly identify those students by clicking the histogram bar for the shorter students. The corresponding rows in the data table are selected. See Figure 1.12.

Figure 1.12 Connection Between Platform Results and Data Table

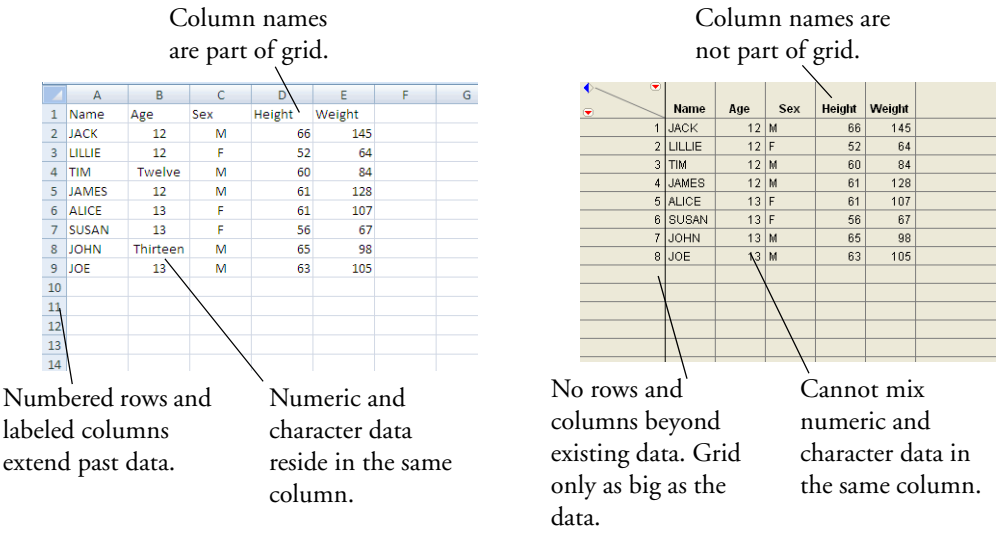


How is JMP Different from Excel?

If you are not familiar with JMP, and are familiar with Microsoft Excel or other spreadsheet applications, there are a number of differences that you should understand. For details of those differences, see Figure 1.13 to Figure 1.16.

Another difference not shown in the figures is that Excel formulas are applied to individual cells. In JMP, formulas are applied only to entire columns.

Figure 1.13 Columns and Data in Excel (left) and JMP (right)



	A	B	C	D	E
1	Name	Age	Sex	Height	Weight
2	JACK	12	M	66	145
3	LILLIE	12	F	52	64
4	TIM	Twelve	M	60	84
5	JAMES	12	M	61	128
6	ALICE	13	F	61	107
7	SUSAN	13	F	56	67
8	JOHN	Thirteen	M	65	98
9	JOE	13	M	63	105

	Name	Age	Sex	Height	Weight
1	JACK	12	M	66	145
2	LILLIE	12	F	52	64
3	TIM	12	M	60	84
4	JAMES	12	M	61	128
5	ALICE	13	F	61	107
6	SUSAN	13	F	56	67
7	JOHN	13	M	65	98
8	JOE	13	M	63	105

Figure 1.14 JMP Multiple Tables

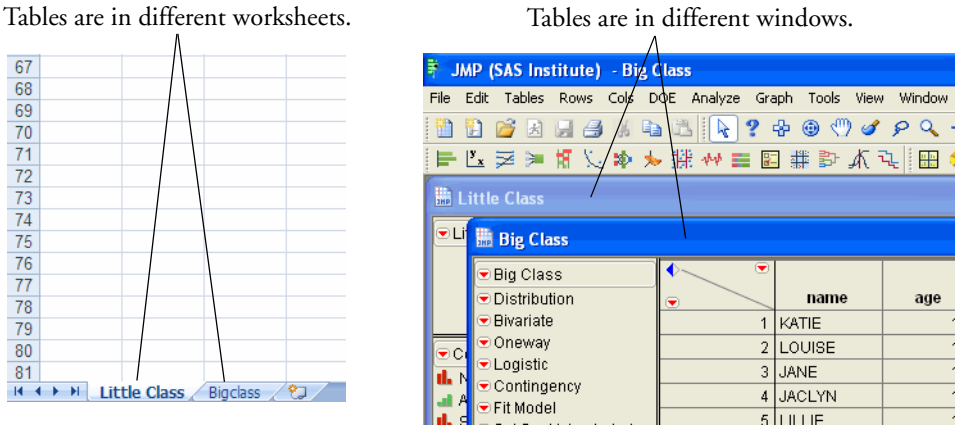


Figure 1.15 Data Placement in Grid

Data can be placed anywhere in grid. For this example, the data starts in row 5 and column 2.

Data always starts in row 1 and column 1.

	A	B	C	D	E	F
1						
2						
3						
4		name	age	sex	height	weight
5		KATIE	12	F	59	95
6		LOUISE	12	F	61	123
7		JANE	12	F	55	74
8		JACLYN	12	F	66	145
9		LILLIE	12	F	52	64

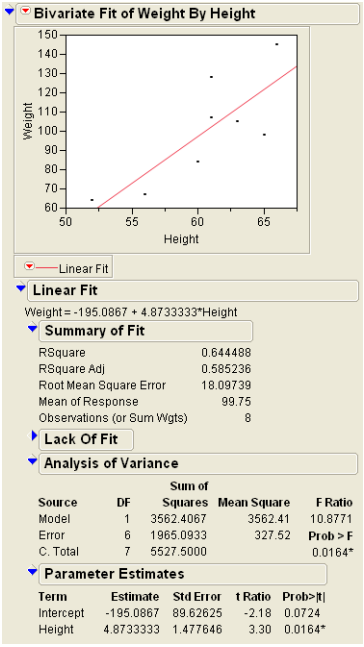
	Name	Age	Sex	Height	Weight
1	JACK	12	M	66	145
2	LILLIE	12	F	52	64
3	TIM	12	M	60	84
4	JAMES	12	M	61	128
5	ALICE	13	F	61	107
6	SUSAN	13	F	56	67
7	JOHN	13	M	65	98
8	JOE	13	M	63	105

Figure 1.16 Results in Excel versus JMP

Results appear in the grid.

10							
11	SUMMARY OUTPUT						
12							
13	*Regression Statistics*						
14	Multiple R	0.802800015					
15	R Square	0.644487864					
16	Adjusted R Square	0.585235841					
17	Standard Error	18.09739085					
18	Observations	8					
19							
20	ANOVA						
21		*df*	*SS*	*MS*	*F*	*Significance F*	
22	Regression	1	3562.406667	3562.407	10.87706	0.016448038	
23	Residual	6	1965.093333	327.5156			
24	Total	7	5527.5				
25							
26		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
27	Intercept	-195.0866667	89.62625095	-2.17667	0.072393	-414.3942019	24.22086853
28	X Variable 1	4.873333333	1.477645775	3.298039	0.016448	1.257664381	8.489002286
29							

Results appear in a separate window.



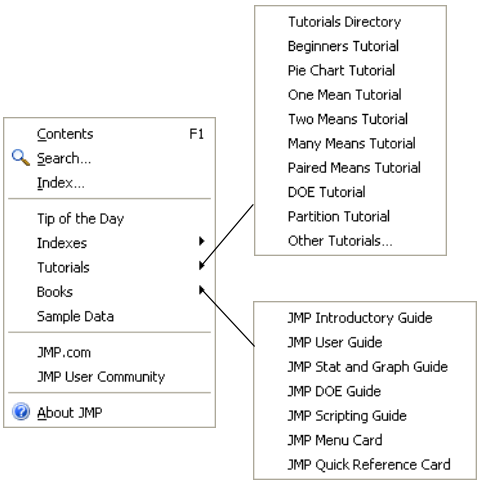
How Do I Access Help?

As you start using JMP, a variety of resources are available to supplement your learning. You have access to JMP's built-in Help menu and various online resources.

Help Menu

From the Help menu on JMP's menu bar, you can access the online documentation, online Help, web information, sample data, and more. See Figure 1.17.

Figure 1.17 Help Menu



Contents, Search, and Index These three menu items access the JMP Help system. The Help system provides navigable and searchable documentation.

Tip of the Day The Tip of the Day is a collection of helpful tips and hints that enhance your experience with JMP.

Indexes The Indexes contain references that give definitions of JMP statistical terms, JSL functions, and JSL objects. For an introduction to JSL scripting, see the Customizing and Extending chapter.

Tutorials Tutorials are self-contained mini lessons that demonstrate how to use some of JMP's statistical and graphical features.

Tip: If you are not familiar with JMP, start with the Beginners Tutorial. Every new JMP user should take this five-minute tutorial on the user-interface basics of JMP.

Books This menu contains links to the full documentation book set, a description of all the JMP menus, and a quick reference describing keyboard shortcuts.

Sample Data This provides access to all the sample data used in the documentation. The sample data helps with learning JMP.

JMP.com JMP.com takes you to the JMP Web site. For more information about JMP's Web site and other resources, see "JMP Web Site," p. 15.

JMP User Community JMP User Community takes you to the online user forum where you can download JMP files, access the JMP Blog, and discuss JMP topics with other users.

JMP Web Site

JMP's Web site (www.jmp.com) contains a wealth of information and provides links to the JMP Blog, JMP Discussion Forum, JMP User Community, the latest news and events, and file sharing. You can register for free weekly webinars, and access podcasts and other JMP literature. The Web site also

contains videos that demonstrate the newest JMP features. You can also learn about joining one of the many regional JMP User's Groups.

Chapter 2

Preparing Data

Prerequisites for Graphing and Analyzing Data

Before you can work with your data, the data has to be in a JMP data table and in the proper format. This chapter contains the following content:

Figure 2.1 Example of a Data Table

Companies
Notes Selected Data o

Columns (8/0)
- Type
- Size Co
- Sales ($M)
- Profits ($M)
- # Employ
- profit/emp
- Assets
- %profit/sales

Rows

All rows	32
Selected	0
Excluded	0
Hidden	0
Labelled	0

	Type	Size Co	Sales ($M)	Profits ($M)	# Employ	profit/emp
1	Computer	small	855.1	31.0	7523	4120.70
2	Pharmaceut	big	5453.5	859.8	40929	21007.11
3	Computer	small	2153.7	153.0	8200	18658.54
4	Pharmaceut	big	6747.0	1102.2	50816	21690.02
5	Computer	small	5284.0	454.0	12068	37620.15
6	Pharmaceut	big	9422.0	747.0	54100	13807.76
7	Computer	small	2876.1	333.3	9500	35084.21
8	Computer	small	709.3	41.4	5000	8280.00
9	Computer	small	2952.1	-680.4	18000	-37800.0
10	Computer	small	784.7	89.0	4708	18903.99
11	Computer	small	1324.3	-119.7	13740	-8711.79
12	Pharmaceut	medium	4175.6	939.5	28200	33315.60
13	Computer	big	11899.0	829.0	95000	8726.32
14	Computer	small	873.6	79.5	8200	9695.12
15	Pharmaceut	big	9844.0	1082.0	83100	13020.46
16	Pharmaceut	small	969.2	227.4	3418	66530.13
17	Pharmaceut	medium	6698.4	1495.4	34400	43470.93
18	Computer	big	5956.0	412.0	56000	7357.14
19	Pharmaceut	big	5903.7	681.1	42100	16178.15
20	Computer	medium	2959.3	252.8	31404	8049.93
21	Pharmaceut	small	1198.3	86.5	8527	10144.25
22	Computer	small	990.5	20.9	8578	2436.47
23	Pharmaceut	medium	3243.0	471.3	21300	22126.76
24	Computer	small	1382.3	0.3	2900	103.45
25	Computer	small	1014.0	47.7	9100	5241.76
26	Computer	small	1769.2	60.8	10200	5960.78
27	Computer	small	1643.9	118.3	9548	12390.03
28	Computer	big	1096.9	-639.3	82300	-7767.92

Contents

Moving Data into JMP

This section describes three ways you can move data into JMP:

- If you want to type data directly into the data table, see "Typing Data," p. 19.
- If you want to copy and paste data from another application, see "Copying and Pasting Data," p. 23.
- If you want to import data from another application, see "Importing Data," p. 23.

You can also import data into JMP from a database. For more information, see the *JMP User Guide*.

Typing Data

To demonstrate directly typing data into a JMP data table, enter the data shown in Figure 2.2.

Figure 2.2 Blood Pressure Data

	Month	Control	Placebo	300mg	450mg
1	March	165	163	166	168
2	April	162	159	165	163
3	May	164	158	161	153
4	June	162	161	158	151
5	July	166	158	160	148
6	August	163	158	157	150

Scenario

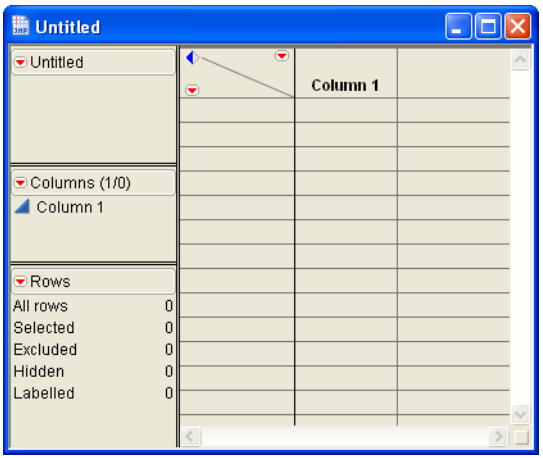
The data are the results from a study that investigated a new blood pressure medication. Each individual's blood pressure was measured over a six month period. Two doses (300mg and 450mg) of the medication were used, along with a control and placebo group. The data is the average blood pressure for each group.

Creating a Data Table

To create the data table shown in Figure 2.2, follow the steps below:

1 Create an empty data table by selecting **File > New > Data Table**. See Figure 2.3.

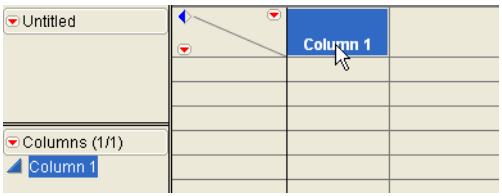
Figure 2.3 Empty Data Table



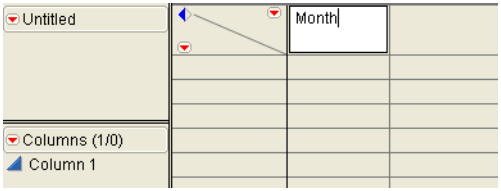
A new data table has one column and no rows.

2 Enter the column name Month. See Figure 2.4.

Figure 2.4 Entering a Column Name



Click once to select



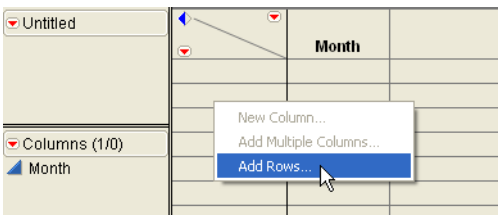
Click again, then type

Note: If you double-click too quickly, the Column Info window appears. You can change the column name there as well.

3 Add rows to the table by following these instructions. See Figure 2.5.

- Right-click in the left margin of the grid.
- Select **Add Rows**.
- Since you want to add six rows, type 6 into the Add Rows window and click **OK**.

Figure 2.5 Adding Rows



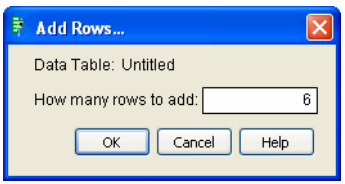
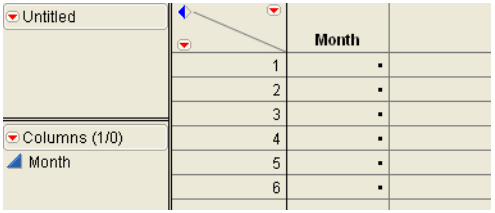


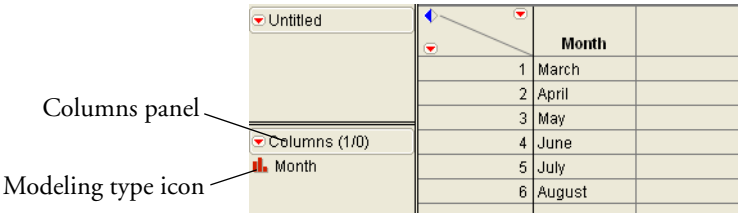
Figure 2.6 shows six rows added to the data table.

Figure 2.6 Six Rows Added



4 Enter the Month information by double-clicking into a cell and typing. The completed column is shown in Figure 2.7.

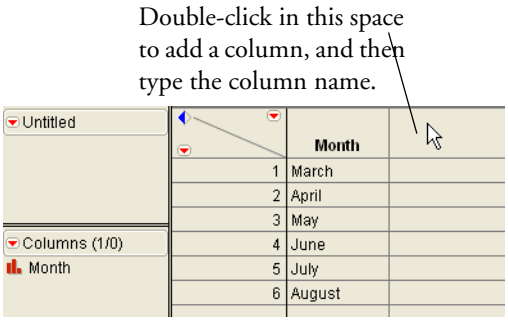
Figure 2.7 Month Column Completed



In the column information panel, look at the modeling type icon to the left of the column name. It has changed to signify that Month is now nominal (previously it was continuous). See Figure 2.6. This difference is important and is discussed in "Viewing or Changing Column Information," p. 33.

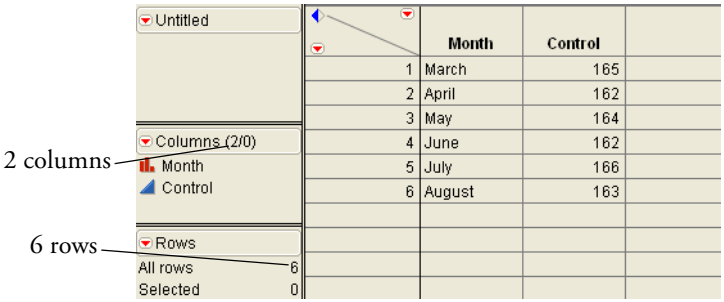
5 Add the Control column. See Figure 2.8 for instructions on adding a new column.

Figure 2.8 Adding a Column



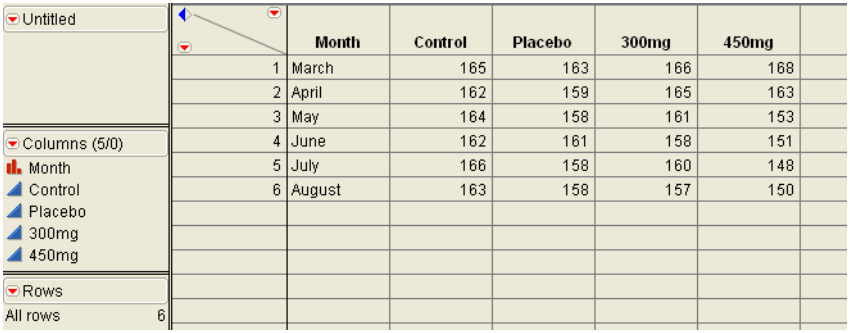
6 Enter the Control data that is shown in Figure 2.2. Your data table now consists of six rows and two columns, as shown in Figure 2.9.

Figure 2.9 Second Column Added



7 Continue adding columns and entering data to create the final data table that has six rows and five columns, as shown in Figure 2.10.

Figure 2.10 Final Data Table



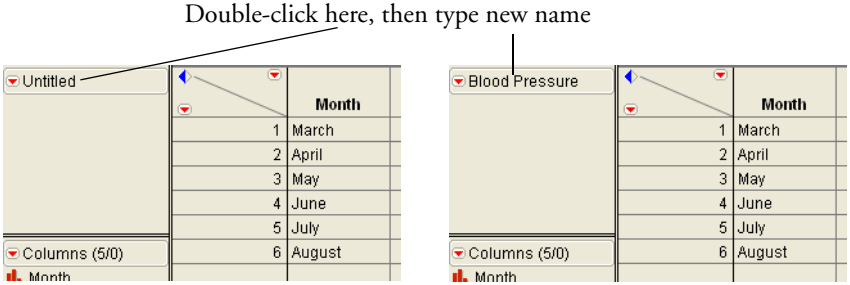
Changing the Data Table Name

You can change the name of the data table by following the steps below:

1 Double-click on the data table name in the Table Panel. See Figure 2.11.

2 Type the new name.

Figure 2.11 Changing the Data Table Name



Copying and Pasting Data

You can move data into JMP by copying and pasting from another application. To demonstrate this approach, use the Bigclass.xls file.

1 Open the Bigclass.xls file, usually found here:
C:\Program Files\SAS\JMP\8\Support Files English\Sample Import Data

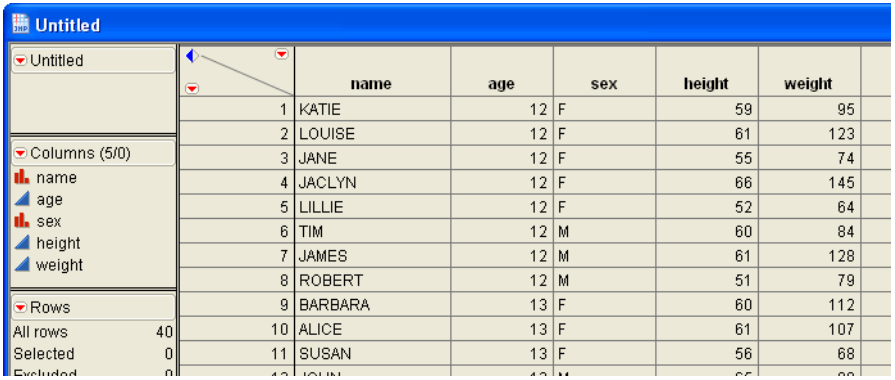
2 Select all of the rows and columns, including the column names. In total, you should select 41 rows and five columns.

3 Copy the selected data.

4 In JMP select **File > New > Data Table** to create an empty table.

5 Select **Edit > Paste with Column Names** to paste the data and column headings. Or, if the data you are pasting into JMP does not have column names, then use **Edit > Paste**. A partial view of the resulting table is shown in Figure 2.12.

Figure 2.12 Copying and Pasting Data



	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112
10	ALICE	13	F	61	107
11	SUSAN	13	F	56	68

Importing Data

You can move data into JMP by importing from another application. To demonstrate importing data into JMP, import from a Microsoft Excel (.xls) file.

1 Select **File > Open**.

2 Navigate to the Sample Import Data folder, usually found here:
C:\Program Files\SAS\JMP\8\Support Files English

3 On the "Files of Type" menu, select **Excel 97-2003 (*.xls)**. See Figure 2.13.

4 Select Bigclass.xls.

5 Click **Open.**

Figure 2.14 shows a partial view of the resulting data table.

Figure 2.13 Importing an Excel File

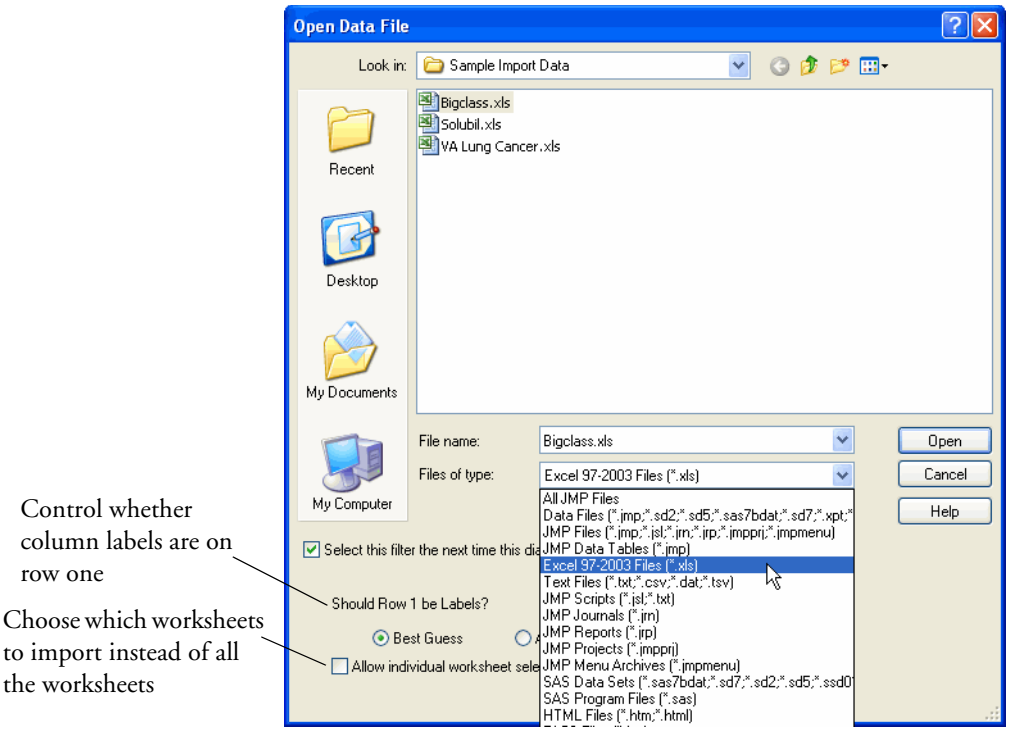


Figure 2.14 Bigclass Data Table

Bigclass

Bigclass
Notes C:\Program Files\SA

Columns (5/0)
name
age
sex
height
weight

Rows
All rows 40
Selected 0
Excluded 0
Hidden 0
Labelled 0

	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	79
9	BARBARA	13	F	60	112
10	ALICE	13	F	61	107
11	SUSAN	13	F	56	68
12	JOHN	13	M	65	98
13	JOE	13	M	63	105
14	MICHAEL	13	M	58	95
15	DAVID	13	M	59	79
16	JUDY	14	F	61	81

The Open Data File window (see Figure 2.13) has two additional options:

- You can control whether JMP assumes that column headings are on row one in the Excel file.
- If your Excel file has multiple worksheets, you can choose which worksheets to import. Each worksheet is imported to a separate data table.

When you import Excel files, JMP predicts where the column headings are and where the data begins. If the column names are not on row one, or if the data does not start on row two, it might be easier to use the copy and paste method described in the section called "Opening Existing JMP Data Tables," p. 25.

Opening Existing JMP Data Tables

To open an existing JMP data table, simply double-click on the file icon. If JMP is not already running, this starts JMP and opens the data table.

You can open an existing JMP file from inside JMP. Select **File > Open**, navigate to the file's location, select the file, and click **Open**.

Accessing the Sample Data Tables

The examples in the JMP documentation use sample data tables. The default location on Windows for the sample data is here:

C:\Program Files\SAS\JMP\8\Support Files English\Sample Data

To quickly access the sample data tables, select **Help > Sample Data**. The Sample Data Index appears. The Sample Data Index groups the data tables by topic. Click a blue triangle to see a list of data tables for that topic. Now click on a link to open a data table.

The Sample Data Index has two buttons:

Open the Sample Data Directory Clicking this button takes you to the sample data directory.

See an Alphabetical List of all Sample Data Files Clicking this button shows an alphabetical list of all the data tables. Click a link to open a data table.

Interacting with Data

This section shows the following:

Editing Data

You often need to edit or change data, either a few cells at a time or for an entire column. This section contains the following information:

The exercises in this section use the Companies.jmp data.

Changing Values

To change a value, double-click in the target cell, and then type the change.

Note: Double-clicking in a cell is not the same as selecting a cell. Double-clicking lets you edit a cell. A single click selects a cell. For more information about selecting rows, columns, and cells, see "Selecting and Deselecting Data," p. 30.

As an example of changing a value, the first row of the Companies.jmp data table contains information about a small computer company. See Figure 2.15.

Figure 2.15 Companies Data Table

	Type	Size Co	Sales ($M)	Profits ($M)	# Employ	profit/emp	Assets	%prof
1	Computer	small	855.1	31.0	7523	4120.70	615.2	
2	Pharmaceut	big	5453.5	859.8	40929	21007.11	4851.6	
3	Computer	small	2153.7	153.0	8200	18658.54	2233.7	

This company's information is incorrect. Correct the information by changing these values:

- Change the Sales ($M) value to $860.
- Change the # Employ value to 7400.
- Change the Assets value to 623.

Recoding Values

You can use recoding to change all the values in a column at once. To demonstrate how to recode values, suppose you are interested in comparing the sales of computer and pharmaceutical companies. When the results of your analysis are displayed, you want the company labels to be Comp and Pharm, not Computer and Pharmaceutical. One option is to go through all 32 rows of data and change all the "Computer" values to Comp and all the "Pharmaceutical" values to Pharm. That process would be tedious and inefficient with hundreds of rows of data. A better option is to use the recode capability:

1 Select the Type column by clicking once on the column heading.
2 Select **Cols > Recode**.
3 In the Recode window, enter the desired values in the **New Values** boxes. For this example, enter Comp in the Computer row, and Pharm in the Pharmaceutical row. See Figure 2.16.
4 Choose the **In Place** option from the menu. See the *JMP User Guide* for explanations of the menu options.
5 Click **OK**.

6 Click **Continue** on the alert window. The values update. See Figure 2.17.

Figure 2.16 Recode Window

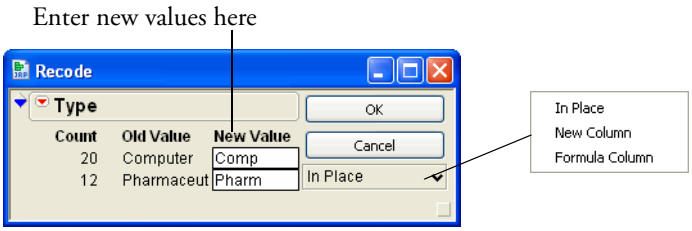
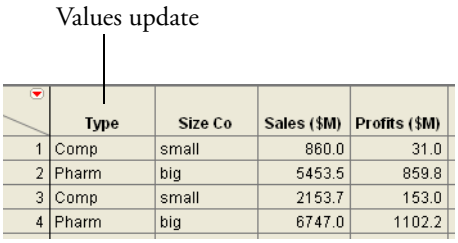


Figure 2.17 Recoded Values



	Type	Size Co	Sales ($M)	Profits ($M)
1	Comp	small	860.0	31.0
2	Pharm	big	5453.5	859.8
3	Comp	small	2153.7	153.0
4	Pharm	big	6747.0	1102.2

Tip: The Value Label column property lets you specify labels to use on output without changing the underlying value. See the *JMP User Guide* for more details.

Searching for Values

In a data table that has thousands or tens of thousands of rows, it can be difficult to locate a particular row just by scrolling through the table. If you are looking for specific information, you can use the search feature to find it. If it finds data that matches the search criteria, it automatically navigates to the first matching row. For example, the Companies data table contains information about a company that has sales of $11899. You can use the search feature to find that row.

1 Select **Edit > Search > Find** to launch the Search window.

2 In the **Find what** box, enter 11899 as shown in Figure 2.18.

3 Click **Find**. JMP finds the first cell that has 11899 in it, and highlights it. See Figure 2.19.

If multiple cells meet the search criteria, you can click **Find** again to highlight the next cell that meets the criteria.

Figure 2.18 Search Window

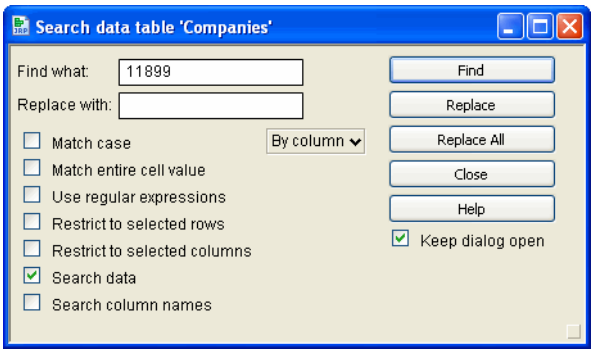


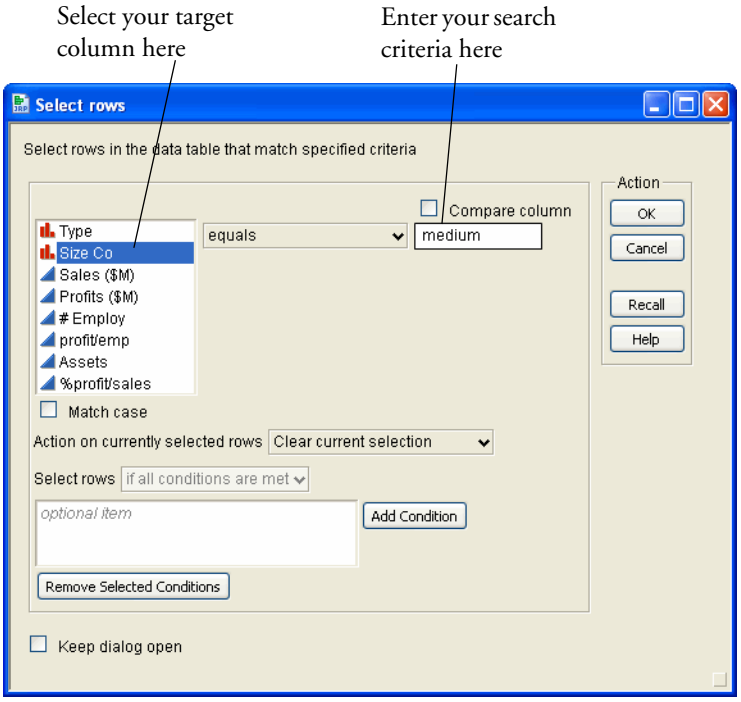
Figure 2.19 Highlighted Search Results

11	Computer	small	1324.3	-119.7	13740
12	Pharmaceut	medium	4175.6	939.5	28200
13	Computer	big	11899.0	829.0	95000
14	Computer	small	873.6	79.5	8200
15	Pharmaceut	big	9844.0	1082.0	83100
16	Pharmaceut	small	969.2	227.4	3418

You can also search for multiple rows at once, with each row matching some criteria. For example, suppose you want to select all of the rows that correspond to medium companies. To do this, follow the steps below:

1 Select **Rows > Row Selection > Select Where** to launch the **Select rows** window.
2 In the column list box on the left, select Size Co.
3 In the text box on the right, enter medium. See Figure 2.20.
4 Click **OK**.

Figure 2.20 Select Rows Window



JMP highlights all of the rows that have Size Co equal to medium. There are seven. See Figure 2.21.

Figure 2.21 Search Results Selected

11	Computer	small	1324.3	-119.7	13740	-8711.79	1040.2	-9.04
12	Pharmaceut	medium	4175.6	939.5	28200	33315.60	5848.0	22.50
13	Computer	big	11899.0	829.0	95000	8726.32	10075.0	6.97
14	Computer	small	873.6	79.5	8200	9695.12	808.0	9.10
15	Pharmaceut	big	9844.0	1082.0	83100	13020.46	7919.0	10.99
16	Pharmaceut	small	969.2	227.4	3418	66530.13	784.0	23.46
17	Pharmaceut	medium	6698.4	1495.4	34400	43470.93	6756.7	22.32
18	Computer	big	5956.0	412.0	56000	7357.14	4500.0	6.92
19	Pharmaceut	big	5903.7	681.1	42100	16178.15	8324.8	11.54
20	Computer	medium	2959.3	252.8	31404	8049.93	5611.1	8.54
21	Pharmaceut	small	1198.3	86.5	8527	10144.25	1791.7	7.22
22	Computer	small	990.5	20.9	8578	2436.47	624.3	2.11
23	Pharmaceut	medium	3243.0	471.3	21300	22126.76	3613.5	14.53
24	Computer	small	1382.3	0.3	2900	103.45	1076.8	0.02
25	Computer	small	1014.0	47.7	9100	5241.76	977.0	4.70
26	Computer	small	1769.2	60.8	10200	5960.78	1269.1	3.44
27	Computer	small	1643.9	118.3	9548	12390.03	1618.8	7.20
28	Computer	big	1096.9	-639.3	82300	-7767.92	10751.0	-58.28
29	Pharmaceut	medium	2916.3	176.0	20100	8756.22	3246.9	6.04
30	Computer	medium	3078.4	-424.3	28334	-14974.9	2725.7	-13.78
31	Pharmaceut	medium	4272.0	412.7	33000	12506.06	3051.6	9.66
32	Computer	big	63438.0	3758.0	383220	9806.38	77734.0	5.92

Selecting and Deselecting Data

You might need to select rows, columns, or cells within a data table. For example, to create a subset table, the corresponding parts of the table must first be selected. Also, selecting rows is helpful in making data points stand out on a graph.

Rows

Table 2.1 describes how to select and deselect rows.

Table 2.1 Selecting and Deselecting Rows

Task	Action
Select rows one at a time	Click on the row number (see Figure 2.22)
Select multiple adjacent rows	Click and drag on the row numbers or Select the beginning row, then hold down the Shift key and click the last row number
Select multiple non-adjacent rows	Select the first row, then hold down the Ctrl key and click the other row numbers
Deselect rows one at a time	Hold down the Ctrl key and click the row numbers
Deselect all rows	Click in the lower-triangular space in the top left corner of the table (see Figure 2.23)

Figure 2.22 Select a Row

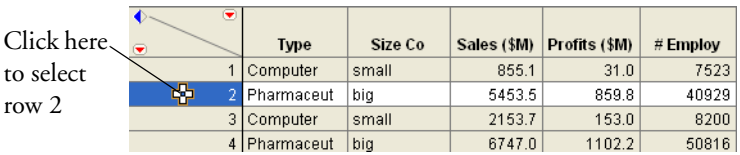
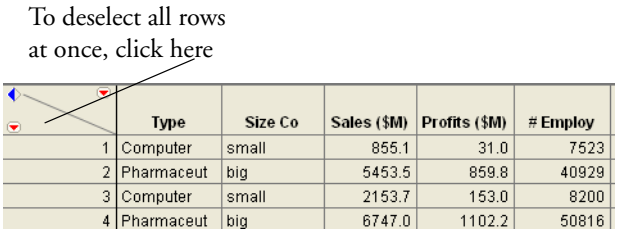


Figure 2.23 Deselecting Rows



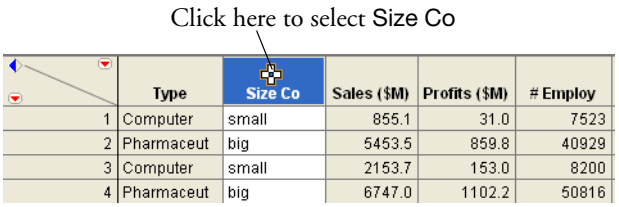
Columns

Table 2.2 describes how to select and deselect columns.

Table 2.2 Selecting and Deselecting Columns

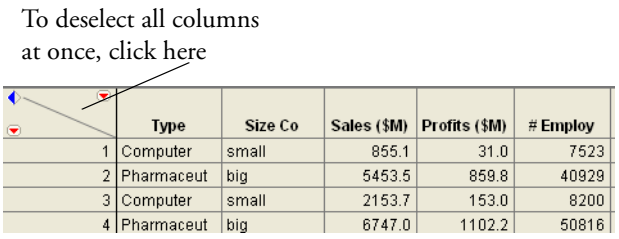
Task	Action
Select columns one at a time	Click the column heading. See Figure 2.24.
Select multiple adjacent columns	Click and drag across the column headings or Select the beginning column, then hold down the Shift key and click the last header
Select multiple non-adjacent columns	Select the first column, then hold down the Ctrl key and click the other column headings
Deselect columns one at a time	Hold down the Ctrl key and click the column heading
Deselect all columns	Click in the upper-triangular space in the top left corner of the table. See Figure 2.25.

Figure 2.24 Selecting a Column



	Type	Size Co	Sales ($M)	Profits ($M)	# Employ
1	Computer	small	855.1	31.0	7523
2	Pharmaceut	big	5453.5	859.8	40929
3	Computer	small	2153.7	153.0	8200
4	Pharmaceut	big	6747.0	1102.2	50816

Figure 2.25 Deselect Columns



	Type	Size Co	Sales ($M)	Profits ($M)	# Employ
1	Computer	small	855.1	31.0	7523
2	Pharmaceut	big	5453.5	859.8	40929
3	Computer	small	2153.7	153.0	8200
4	Pharmaceut	big	6747.0	1102.2	50816

Cells

Table 2.3 describes how to select and deselect cells.

Table 2.3 Selecting and Deselecting Cells

Task	Action
Select cells one at a time	Click in the cell
Select multiple adjacent cells	Click and drag across the cells or Select the beginning cell, then hold down the Shift key and click the last cell
Select multiple non-adjacent cells	Select the first cell, then hold down the Ctrl key and click the other cells
Deselect all cells	Click in the upper and lower triangular spaces in the top left corner of the table

Creating Patterned Data

You can populate a column with patterned data using the Fill features. The Fill feature is especially useful if your data table is large, and typing in the values for each row would be cumbersome.

For example, if you want to fill a column with the pattern 1,2,3,1,2,3,.... follow the steps below:

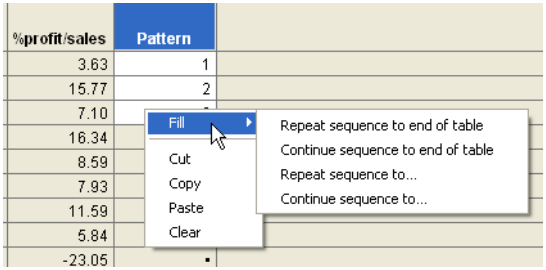
1 Typing one instance of the pattern in the top rows, as shown in Figure 2.26.

Figure 2.26 One Instance of the Pattern

%profit/sales	Pattern
3.63	1
15.77	2
7.10	3
16.34	▪
8.59	▪
7.93	▪

2 Select the three cells, and right-click anywhere in the selected cells. See Figure 2.27.

Figure 2.27 Fill Submenu



3 Select **Repeat sequence to end of table**. This command repeats the pattern to the end of the table.

If, instead of repeating a pattern (1,2,3,1,2,3,...), you want to continue a pattern (1,2,3,4,5,6,...), select **Continue sequence to end of table**. This command can also be used to generate patterns like 1,1,1,2,2,2,3,3,3,....

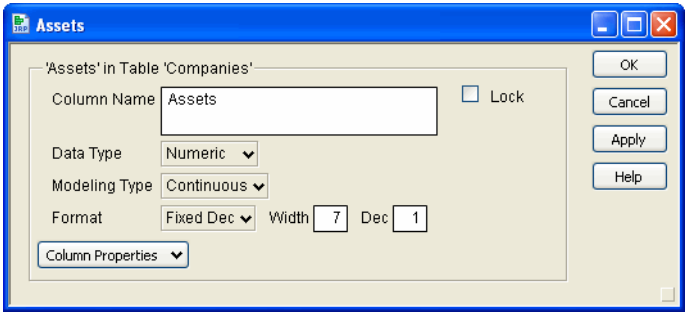
The Fill options can recognize simple arithmetic and geometric sequences. For character data, the Fill features are able to repeat the values only.

Viewing or Changing Column Information

Information about a column is not limited to the data in the column. Other information, such as data type, modeling type, format, and formulas are also important to understand. For example, JMP performs different analyses depending on the modeling type you assign to a column.

To view or change column characteristics, double-click on the column heading. You can also right-click on the column heading and select **Column Info...**. The Column Info window appears. See Figure 2.28.

Figure 2.28 Column Info Window



The Column Info window contains these features:

Column Name You can enter or change the column name by clicking in this box. The text in the box reflects what appears in the column heading. No two columns can have the same column name.

Data Type Click here to choose the data type. The options are:

Numeric specifies the column values as numbers.

Character specifies the column values as non-numeric, such as letters or symbols.

Row State specifies the column values as row states. See the *JMP User Guide* for more details.

Modeling Type Click here to choose the modeling type. This tells JMP how to treat the values in an analysis. The options are:

Continuous means that the values are numeric only.

Ordinal means that the values are either numeric or character, and are ordered categories.

Nominal means that the values are either numeric or character, but not ordered.

Format Click here to specify the format of numeric values (this option is not available for character data). Here are some of the common formats:

Best lets JMP choose the best display format.

Fixed Dec lets you specify the number of displayed decimal places.

Date lets you choose the display and input syntax for dates.

Time is used to choose the display syntax for time values. You can also choose the input syntax.

Currency lets you choose the display syntax for currency values.

Column Properties You can set special column properties such as formulas, notes, and value orders (for details see the *JMP User Guide*).

Lock Lets you lock the column, so the values in the column cannot be changed.

Manipulating Data

You can use the commands on the **Tables** menu to summarize and manipulate data. This section discusses five of the commands:

- Summary
- Tabulate
- Subset
- Join
- Sort

For complete details about these and the other Tables menu commands, see the *JMP User Guide*.

Requesting Summary Statistics

You can create summary tables by using either the Summary or Tabulate menu items on the Tables menu.

Summary

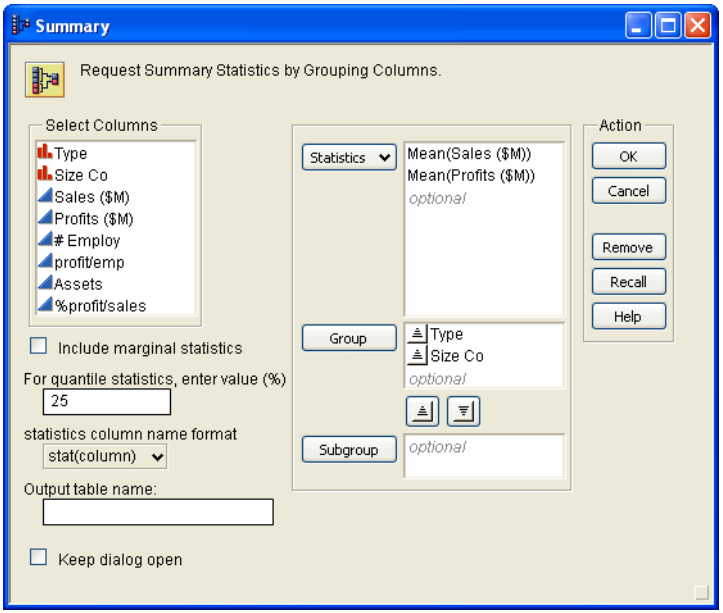
A table summary provides statistics for each level of a grouping variable. As an example, you can look at financial data for computer and pharmaceutical companies. Suppose you want to calculate the mean of sales and the mean of profits, for each combination of company type and size.

1 Open the Companies.jmp sample data table.
2 Select **Tables > Summary**.
3 Assign Type to the Group role by performing one of the following actions:
   – Select Type in the Select Columns box, and click **Group**.
   – Click and drag Type to the Group box.
4 Assign Size Co to the Group role using the same process.
5 Include the mean of Sales ($M) and the mean of Profits ($M) in the Statistics box:
   – Select Sales ($M) in the Select Columns box and click **Statistics > Mean**.

– Select Profits ($M) in the Select Columns box and click **Statistics > Mean**. The completed Summary window is shown in Figure 2.29.

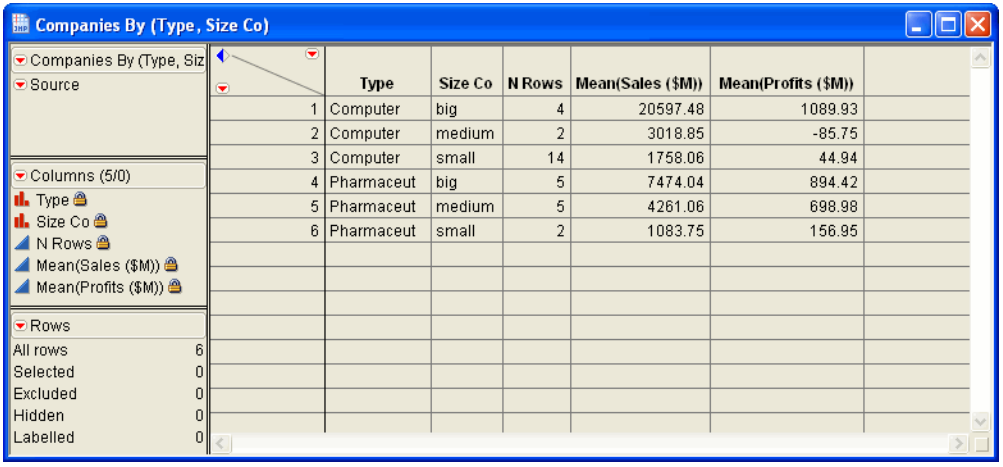
6 Click **OK**.

Figure 2.29 Summary Window



JMP calculates the mean of Sales ($M) and the mean of Profit ($M) for each combination of Type and Size Co. The resulting summary table is shown in Figure 2.30.

Figure 2.30 Summary Table



	Type	Size Co	N Rows	Mean(Sales ($M))	Mean(Profits ($M))
1	Computer	big	4	20597.48	1089.93
2	Computer	medium	2	3018.85	-85.75
3	Computer	small	14	1758.06	44.94
4	Pharmaceut	big	5	7474.04	894.42
5	Pharmaceut	medium	5	4261.06	698.98
6	Pharmaceut	small	2	1083.75	156.95

The summary table contains the following:

- There are columns for each grouping variable (in this example Type and Size Co).

- The N Rows column shows the number of rows from the original table corresponding to that combination of grouping variables. For example, the original data table contains 14 rows corresponding to small computer companies.
- There is a column for each summary statistic requested. In this example, there is a column for the mean of Sales ($M) and one for the mean of Profits ($M).

The summary table is linked to the source table. Click on a row in the summary table to select the corresponding rows in the source table.

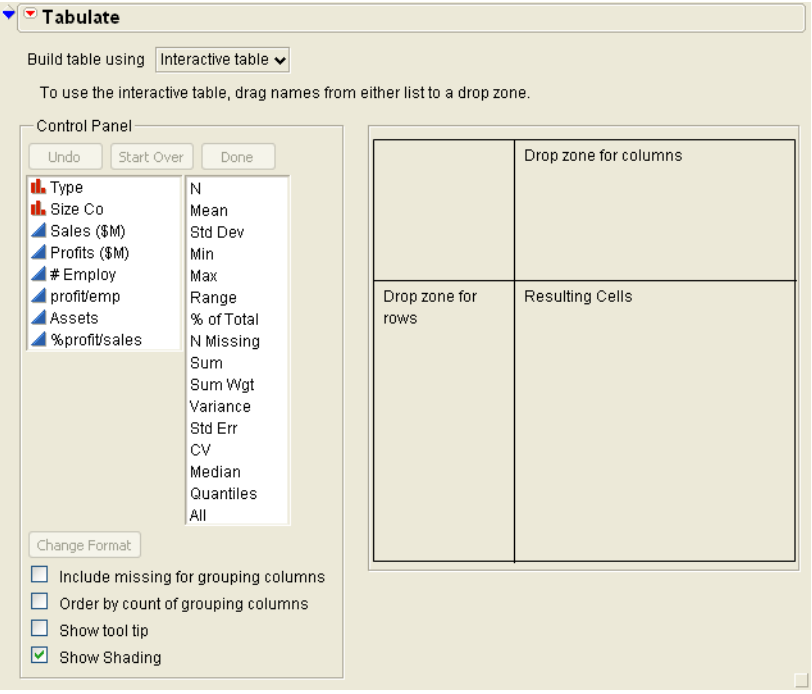
Tabulate

Using the Tabulate menu, you can drag columns into a workspace, creating summary statistics for each combination of grouping variables. As an example, you can look at financial data for computer and pharmaceutical companies. Suppose you want to calculate the mean of sales and the mean of profits, for each combination of company type and size.

1 Open the Companies.jmp sample data table.

2 Select **Tables** > **Tabulate**. Figure 2.31 shows the workspace.

Figure 2.31 Tabulate Workspace



3 Select both Type and Size Co.

4 Drag them to the **Drop zone for rows**. See Figure 2.32. Select **Add Grouping Columns**.

The tabulation gives the number of rows per group. See Figure 2.33.

Figure 2.32 Assign columns to Row Zone

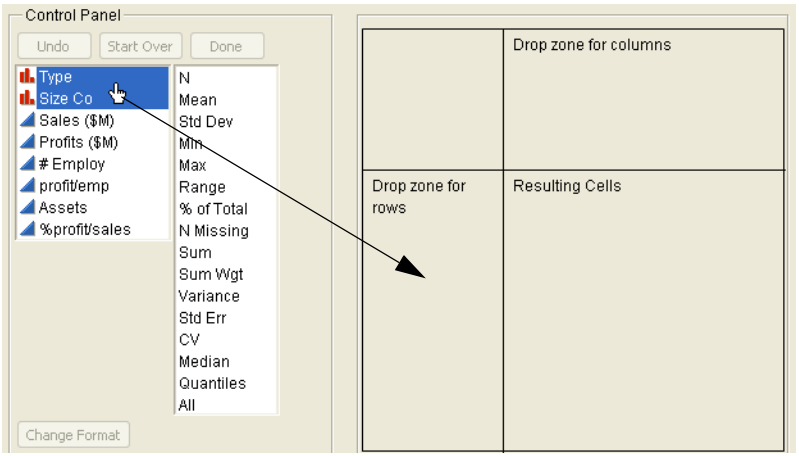


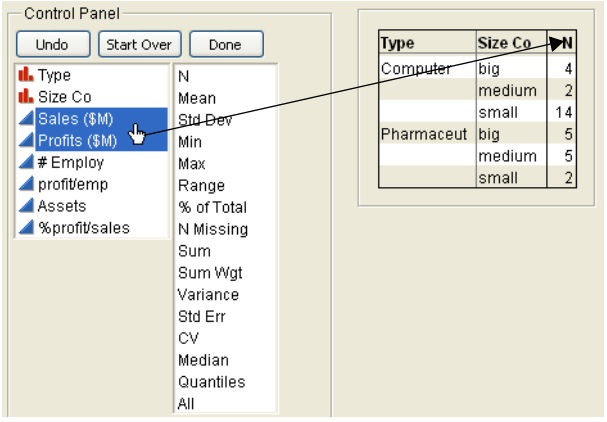
Figure 2.33 Initial Tabulation

Type	Size Co	N
Computer	big	4
	medium	2
	small	14
Pharmaceut	big	5
	medium	5
	small	2

5 Select both Sales ($M) and Profits ($M).

6 Drag them over the **N** in the table. See Figure 2.34. Select **Add Analysis Columns**.

Figure 2.34 Adding Sales and Profit



The tabulation gives the sum of Sales ($M) and the sum of Profits ($M) per group. See Figure 2.35. The final step is to change the sums to means.

Figure 2.35 Tabulation of Sums

		Sales ($M)	Profits ($M)
Type	Size Co	Sum	Sum
Computer	big	82389.9	4359.7
	medium	6037.7	-171.5
	small	24612.8	629.1
Pharmaceut	big	37370.2	4472.1
	medium	21305.3	3494.9
	small	2167.5	313.9

7 Right-click on **Sum** and select **Statistics** > **Mean**. The sums are replaced by the means for each group. See Figure 2.36.

Figure 2.36 Final Tabulation

		Sales ($M)	Profits ($M)
Type	Size Co	Mean	Mean
Computer	big	20597.48	1089.9
	medium	3018.85	-85.75
	small	1758.06	44.94
Pharmaceut	big	7474.04	894.42
	medium	4261.06	698.98
	small	1083.75	156.95

The means are the same as those obtained using **Tables** > **Summary**. See Figure 2.30.

Creating Subsets

You can create a subset of data by extracting portions of a data table into a new table. This can be helpful if you have a large data table, and are interested in examining rows that are scattered randomly throughout the table. It is easier to examine all those rows together in one place and in a smaller table.

Creating a subset is a two-step process. You select the target data, and then extract the data into a new table.

As an example, you can look at financial data for big, medium, and small companies. Suppose you are interested in creating a data table that has only the medium companies. In other words, you want to create a subset of the main table, consisting only of those rows for which Size Co is medium. Also, you want only the Sales ($M), Profits ($M), and Type columns in the subset table.

To create the subset, follow these steps:

1 Open the Companies.jmp sample data table.
2 Select the rows that correspond to males:
   – Select **Rows > Row Selection > Select Where**.
   – Select Size Co in the column list box on the left.
   – Enter medium in the text enter box.
   – Click **OK**.
3 Select the Type, Sales ($M), and Profits ($M) columns.
4 Extract the selected cells into a new table:
   – Select **Tables > Subset** to launch the Subset window. See Figure 2.37.

– Click **OK**.

The resulting data table has 7 rows and 3 columns. See Figure 2.38. For complete details about the Subset window, see the *JMP User Guide*.

Figure 2.37 Subset Window

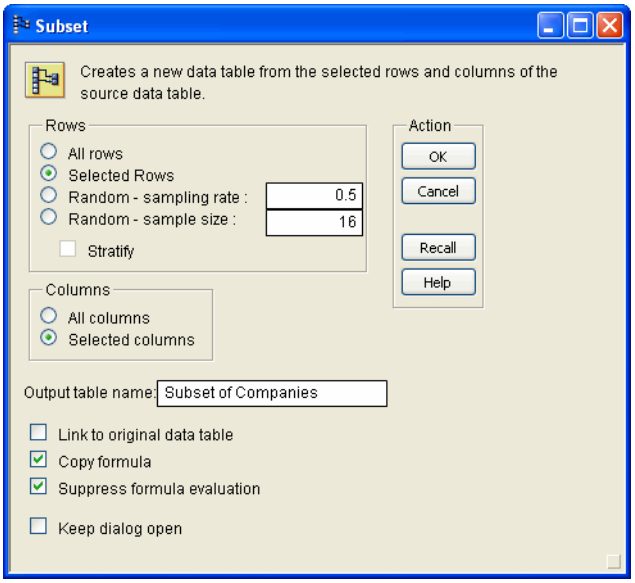
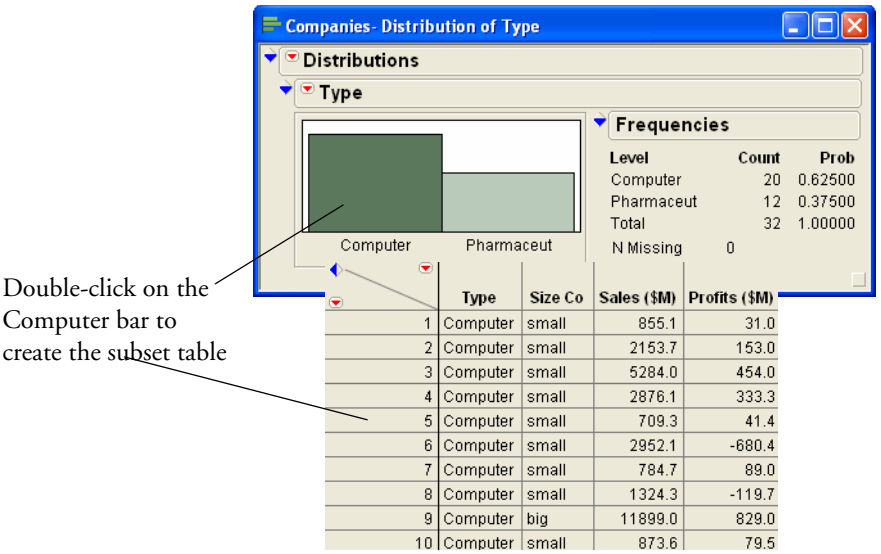


Figure 2.38 Resulting Subset Table

	Type	Sales ($M)	Profits ($M)
1	Pharmaceut	4175.6	939.5
2	Pharmaceut	6698.4	1495.4
3	Computer	2959.3	252.8
4	Pharmaceut	3243.0	471.3
5	Pharmaceut	2916.3	176.0
6	Computer	3078.4	-424.3
7	Pharmaceut	4272.0	412.7

Another way to create simple subsets uses the connection between platform results and data tables. For example, to create a subset of the Computer companies using Companies.jmp, first use the **Distribution** platform to create a histogram of the column Type. Then double-click on the histogram bar corresponding to Computer, and a subset table of the Computer companies is created.

Figure 2.39 Double-Click on Bar to Create Subset Table



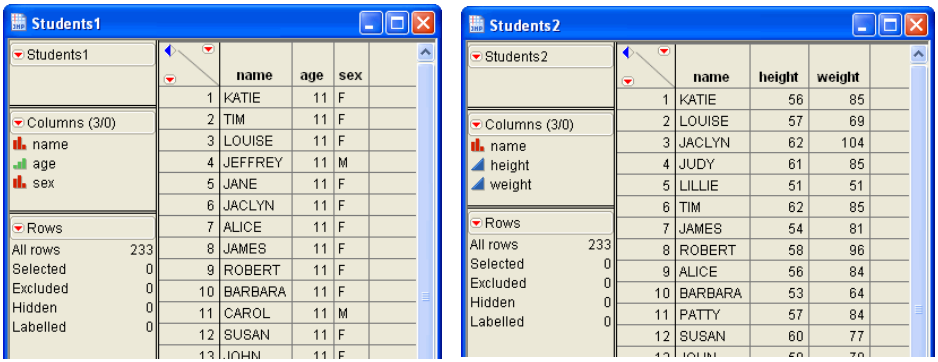
Caution: This method creates a *linked* subset table. This means if you make any changes to the data in the subset table, the corresponding value changes in the source table.

Joining Data Tables

You can combine information from multiple data tables into a single data table. For example, suppose you have a data table that contains students' names, ages, and sexes. In another table, you have the students' names, heights, and weights. To perform an analysis involving age and height, you need to have the data in the same table. Also, the students are not listed in the same order in both tables, making a simple copy and paste impossible.

The sample data tables for this example are called Students1.jmp and Students2.jmp. See Figure 2.40. The columns of interest are age and height, but they are in different tables. Both tables have the column name, and that column can be used to join the tables into one.

Figure 2.40 Tables to be Joined



To join the tables, follow these steps:

1 Open the Students1.jmp and Students2.jmp data tables.
2 Click on Students1.jmp making it the active table.
3 Select **Tables > Join**.
4 In the **Join...With** box select Students2.
5 In the **Matching Specification** panel, select **By Matching Columns** if it is not already selected.
6 In the **Source Columns** boxes, select name in both boxes, and then click **Match**.
7 Both tables have duplicate rows, and you do not want to include those, so check **Drop Multiples** for both Main Table and With Table. Figure 2.41 shows the completed window.
8 Click **OK**.

Figure 2.41 Completed Join Window

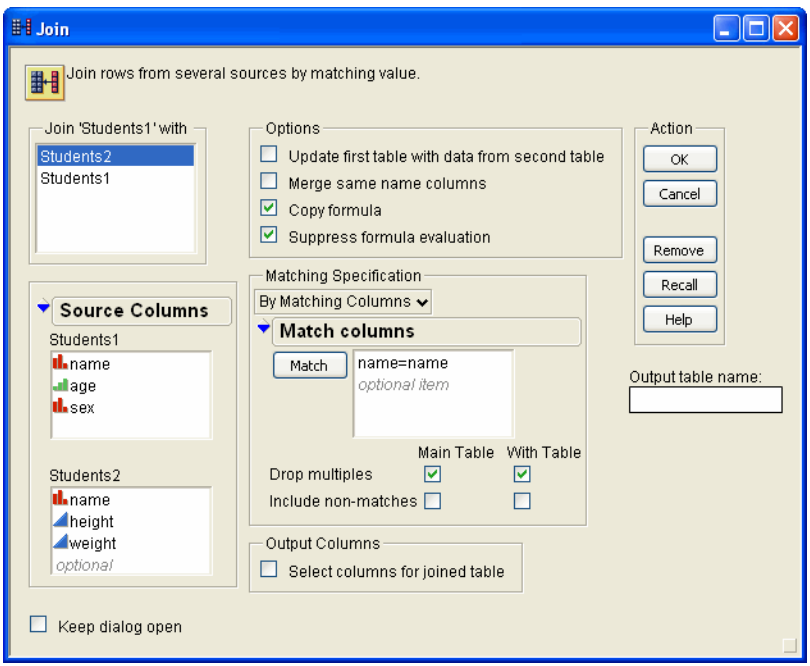


Figure 2.42 shows the new joined table.

Figure 2.42 Joined Table

Untitled 12
Source

Columns (6/0)
name of Students1
age
sex
name of Students2
height
weight

Rows
All rows 39
Selected 0
Excluded 0
Hidden 0
Labelled 0

	name of Students1	age	sex	name of Students2	height	weight
1	ALFRED	11	M	ALFRED	59	87
2	ALICE	11	F	ALICE	56	84
3	AMY	11	M	AMY	55	74
4	BARBARA	11	F	BARBARA	53	64
5	CAROL	11	M	CAROL	60	95
6	CHRIS	11	M	CHRIS	58	85
7	CLAY	12	F	CLAY	63	108
8	DANNY	12	F	DANNY	59	89
9	DAVID	11	F	DAVID	56	73
10	EDWARD	11	M	EDWARD	58	101
11	ELIZABETH	11	M	ELIZABETH	61	87
12	FREDERICK	11	M	FREDERICK	58	75
13	HENRY	11	M	HENRY	55	70
14	JACLYN	11	F	JACLYN	62	104
15	JAMES	11	F	JAMES	54	81

No two columns in a data table can have the same header. Because both original tables contain a column called name, JMP renames those columns in the final table. For example, the column name from the Students1 table becomes name of Students1, and the same for Students2. You can delete one of the name columns, and change the title of the other one back to name.

Tip: The **Select columns for joined table** option on the Join window lets you choose which columns appear in the final table, negating the need for JMP to change the names. See the *JMP User Guide* for details.

Sorting Tables

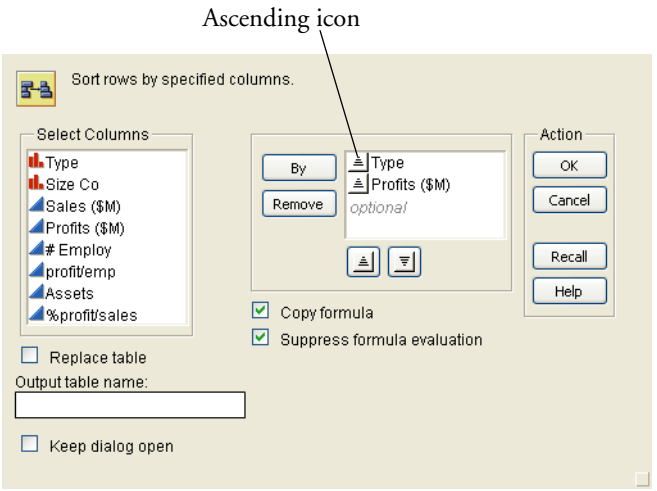
You can sort a data table by one or more columns in the data table. As an example, you can look at financial data for computer and pharmaceutical companies. Suppose you want to sort the data table by Type, then by Profits ($M). Additionally, you want Profits ($M) to be descending within each Type.

To sort the table, do the following:

1 Open the Companies.jmp sample data table.
2 Select **Tables** > **Sort**.
3 To assign Type as a sorting variable, select Type and click **By**.
4 Assign Profits ($M) as a sorting variable by the same process.

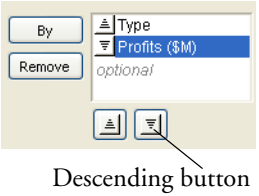
At this point, both variables are set to be sorted in ascending order. See the ascending icon next to the variables in Figure 2.43.

Figure 2.43 Sort Ascending Icons



5 To change Profits ($M) to sort in descending order, select Profits ($M) and click the descending button. See Figure 2.44. The icon next to Profits ($M) changes to descending.

Figure 2.44 Change Profits to Descending



6 Check the **Replace Table** check box. When checked, this option tells JMP to sort the original data table instead of creating a new table with the sorted values. This option is not available if there are any open report windows created from the original data table. The completed sort window is shown in Figure 2.45.

7 Click **OK**. Figure 2.46 shows a portion of the sorted table.

Figure 2.45 Completed Sort Window

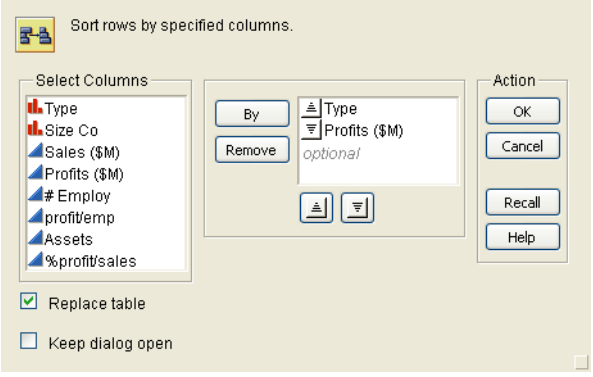


Figure 2.46 Sorted Table

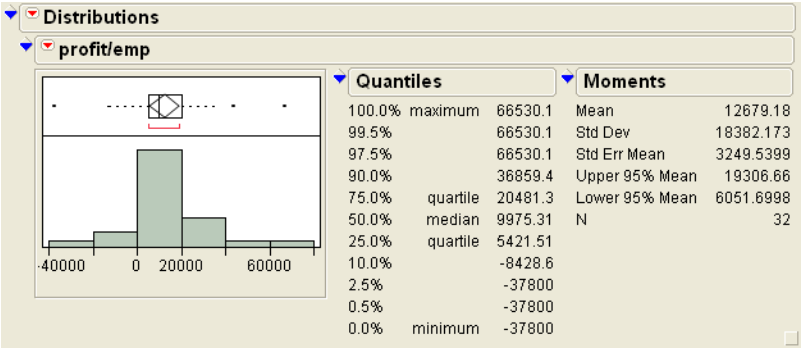
	Type	Size Co	Sales ($M)	Profits ($M)
1	Computer	big	63438.0	3758.0
2	Computer	big	11899.0	829.0
3	Computer	small	5284.0	454.0
4	Computer	big	5956.0	412.0
5	Computer	small	2876.1	333.3
6	Computer	medium	2959.3	252.8
7	Computer	small	2153.7	153.0
8	Computer	small	1643.9	118.3
9	Computer	small	784.7	89.0
10	Computer	small	873.6	79.5
11	Computer	small	1769.2	60.8
12	Computer	small	1014.0	47.7
13	Computer	small	709.3	41.4
14	Computer	small	855.1	31.0
15	Computer	small	990.5	20.9
16	Computer	small	1382.3	0.3
17	Computer	small	1324.3	-119.7
18	Computer	medium	3078.4	-424.3
19	Computer	big	1096.9	-639.3
20	Computer	small	2952.1	-680.4
21	Pharmaceut	medium	6698.4	1495.4
22	Pharmaceut	big	6747.0	1102.2
23	Pharmaceut	big	9844.0	1082.0
24	Pharmaceut	medium	4175.6	939.5

Filtering Data

Using the **Data Filter** commands and options, you can interactively select complex subsets of data, hide these subsets in plots, or exclude them from analyses. As an example, you can look at profit per employee for computer and pharmaceutical companies.

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Distribution**.
3 Assign profit/emp to the **Y, Columns** role.
4 Click **OK**. Figure 2.47 shows the distribution results.

Figure 2.47 Distribution of profit/emp



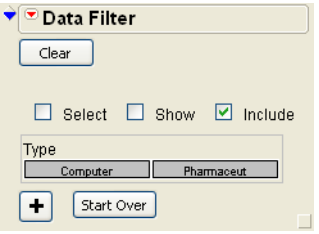
5 Turn on Automatic Recalc by clicking on the red triangle menu next to **Distributions**, and selecting **Script** > **Automatic Recalc**.

6 Select **Rows** > **Data Filter**.

7 Select Type and click **Add**.

8 Uncheck **Select** and check **Include**. Figure 2.48 shows the Data Filter window.

Figure 2.48 Data Filter on Type



9 To filter out the Pharmaceutical companies from the Distribution results, and include only the Computer companies, click the **Computer** box on the Data Filter window. See Figure 2.49.

The distribution results change to only include Computer companies. See Figure 2.50.

Figure 2.49 Click the Computer Box

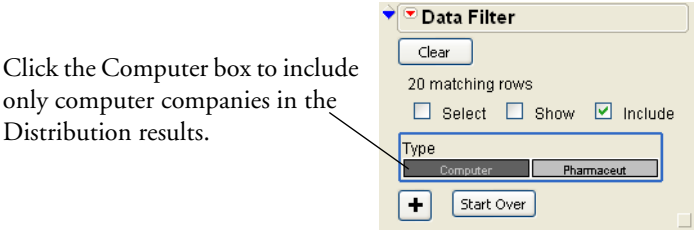
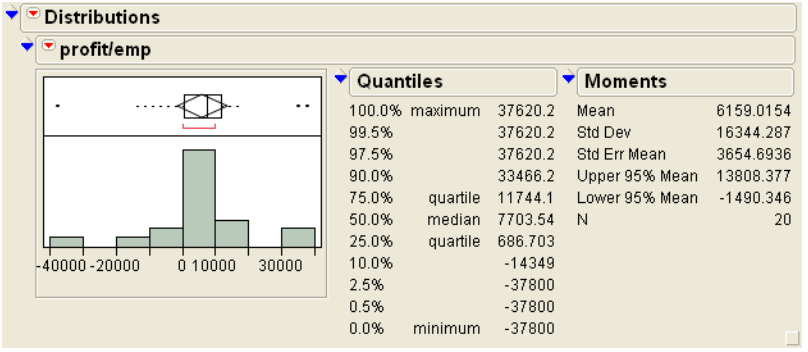


Figure 2.50 Distribution for Computer Companies



To change the Distribution results to include only the Pharmaceutical companies, click the **Pharmaceut** button on the Data Filter window.

Calculating Values Using Formulas

Using the formula editor, you can create columns that contain calculated values.

Scenario

As an example of using the formula editor, use the sample data table On-Time Arrivals.jmp. This data reflects the percent of on-time arrivals for several airlines, measured in March, June, and August of 1999. See Figure 2.51.

Figure 2.51 On-Time Arrivals Data

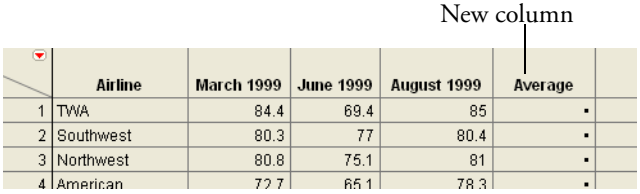
	Airline	March 1999	June 1999	August 1999
1	TWA	84.4	69.4	85
2	Southwest	80.3	77	80.4
3	Northwest	80.8	75.1	81
4	American	72.7	65.1	78.3
5	Delta	78.7	72.2	77.7
6	Continental	79.3	68.4	75.1
7	United	78.6	69.2	71.6
8	US Airways	73.6	68.9	70.1
9	Alaska	71.9	75.4	64.4
10	American West	76.5	70.3	62.5

Creating the Formula

Suppose you want to create a new column that contains the average on-time percent for each airline. Follow these steps:

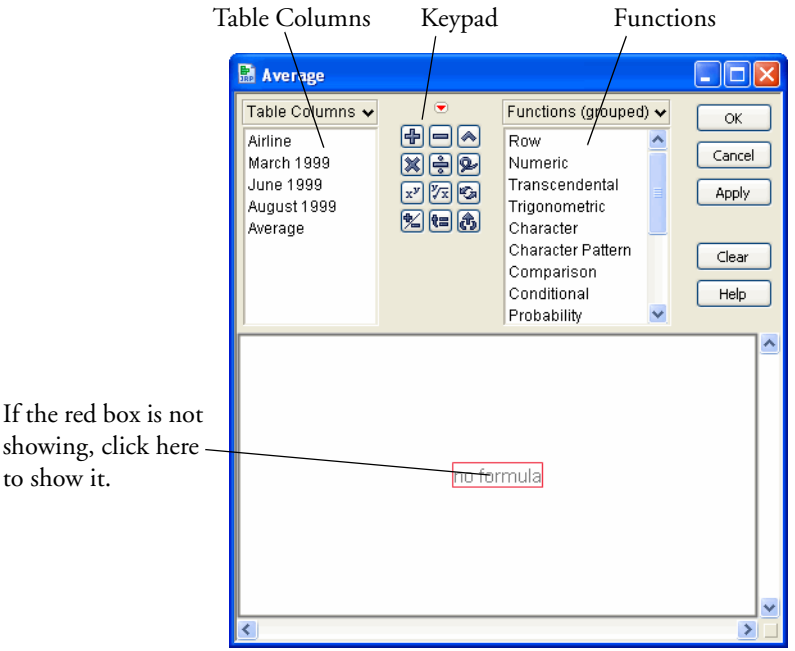
1 Create a new column by double-clicking to the right of the last column. Enter a column name. See Figure 2.52.

Figure 2.52 Creating a New Column



2 Right-click on the column heading of the new column and select **Formula**. The Formula Editor window opens. See Figure 2.53.

Figure 2.53 Formula Editor



Create the formula for the average:

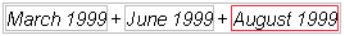
3 From the Table Columns list, select March 1999.

4 Click the ⊞ button.

5 Select June 1999, followed by another ⊞ sign.

6 Select August 1999. At this point, the formula looks like the one shown in Figure 2.54.

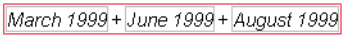
Figure 2.54 Sum of the Months



Notice that only August 1999 is selected (has the red box around it).

7 Click on the box surrounding the entire formula. The red box now surrounds the entire formula as shown in Figure 2.55.

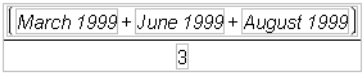
Figure 2.55 Entire Formula Selected



8 Click the ⊟ button.

9 Type a 3 in the denominator box, and then click outside the formula in any of the white space. The finished formula is shown in Figure 2.56.

Figure 2.56 Finished Formula



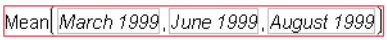
10 Click **OK**. The new column is populated with the averages. See Figure 2.57.

Figure 2.57 New Column Showing Averages

	Airline	March 1999	June 1999	August 1999	Average
1	TWA	84.4	69.4	85	79.6
2	Southwest	80.3	77	80.4	79.2333333
3	Northwest	80.8	75.1	81	78.9666667
4	American	72.7	65.1	78.3	72.0333333
5	Delta	78.7	72.2	77.7	76.2
6	Continental	79.3	68.4	75.1	74.2666667
7	United	78.6	69.2	71.6	73.1333333
8	US Airways	73.6	68.9	70.1	70.8666667
9	Alaska	71.9	75.4	64.4	70.5666667
10	American West	76.5	70.3	62.5	69.7666667

The Formula Editor has many built-in arithmetic and statistical functions. For example, another way to calculate the average on-time arrival percentage is to use the **Mean( )** function on the Statistical functions list. Figure 2.58 shows the Mean( ) function for calculating the average on-time arrival percentage.

Figure 2.58 Mean Function



This is just a basic example of what you can do with the Formula Editor. For details about all the Formula Editor functions, see the *JMP User Guide*.

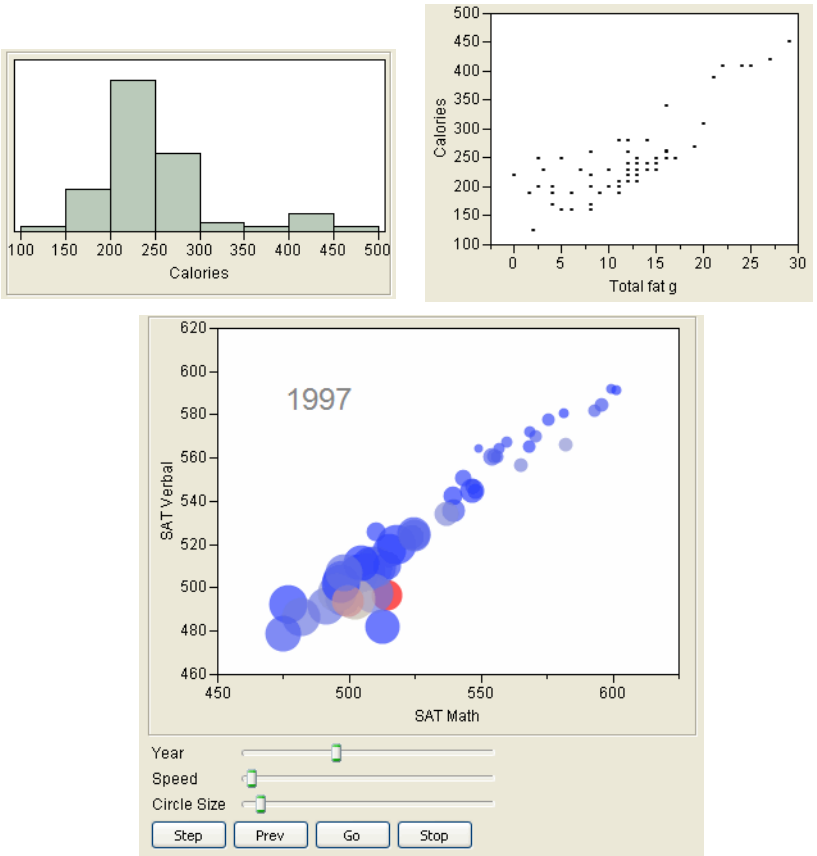
Chapter 3

Visualizing Your Data

Graphing Data

Visualizing your data should always be the first step in any analysis. One of JMP's primary strengths is the ability it gives you to visualize and interact with your data. You do not have to be a statistician to get information and intelligence from your data. Using JMP, you can explore relationships, follow the trends, dig into areas that interest you, and be led in new directions.

Figure 3.1 Visualizing Data with JMP



Contents

About This Chapter

This chapter presents several of the most common graphs and plots that enable you to visualize and explore data in JMP. This chapter is an introduction to JMP's graphical tools and platforms. You can use JMP to visualize the distribution of single variables, or the relationships among multiple variables.

Single variable graphs include the following:

Multiple variable graphs include the following:

Note: This is not an exhaustive list of the graphical capabilities in JMP. For details about JMP's graphical tools and platforms, see the *JMP Statistics and Graphics Guide*.

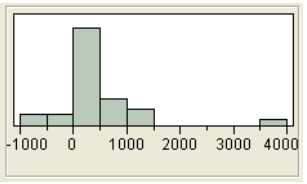
Using Single Variable Graphs

This section covers two graphs you can use to visualize the distribution of a single variable:

You can create both of these graphs using the **Distribution** platform.

Histograms

Figure 3.2 Example of a Histogram



The histogram is one of the most useful graphical tools for understanding the distribution of a continuous variable. Histograms enable you to do the following:

- Quickly assess the average value and variation in your data.
- Find extreme values in your data.

Scenario

To use a histogram, look at sample data for a group of companies. The information includes financial data for each company. For this example, you focus on profits.

Using a histogram, you want to answer these questions:

- Generally, how much profit do the companies earn?
- What is the average profit?
- Are there any outliers?

Creating the Histogram

To answer your questions, create a histogram of Profits ($M):

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Distribution**.
3 Assign Profits ($M) to the **Y, Columns** role. Figure 3.3 shows the completed window.
4 Click **OK**. Figure 3.4 shows the histogram.

Figure 3.3 Distribution Window for Profits ($M)

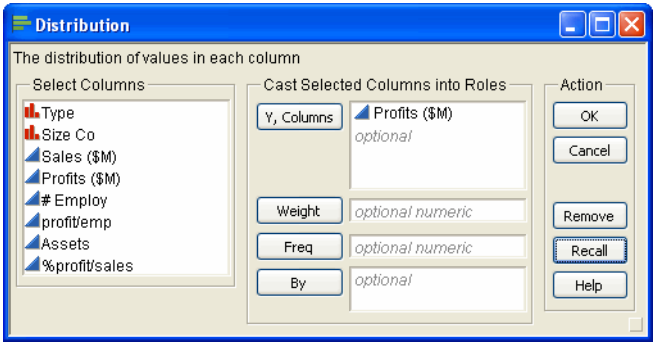
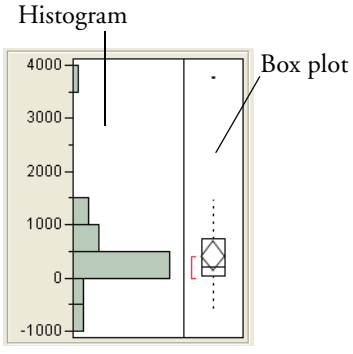


Figure 3.4 Histogram of Profits ($M)



Interpreting the Histogram

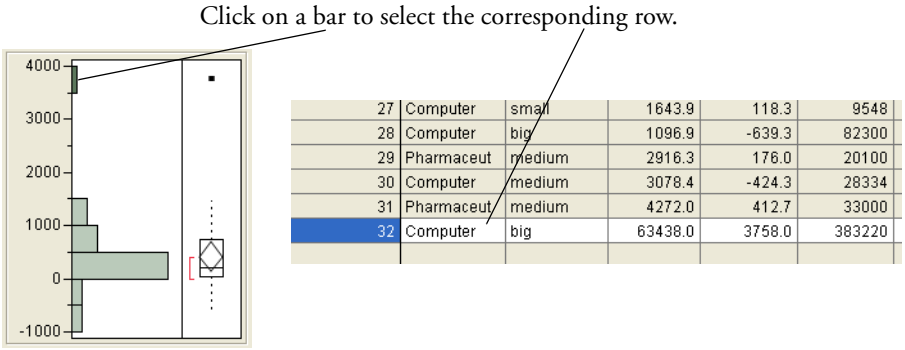
The histogram provides these answers:

- Most companies profits are between $-1000 and $1500.
- The average profit is approximately $500.
- One company has significantly higher profits than the others, and might be an outlier. An outlier is a data point that is separated from the general pattern of the other students.

Interacting with the Histogram

Click on a histogram bar to select the corresponding rows in the data table. See Figure 3.5.

Figure 3.5 Selecting Rows

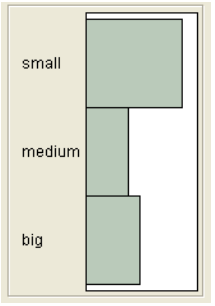


In addition to the histogram, you can also see the following:

- The box plot, which is another graphical summary of the data. For detailed information about the box plot, see the *JMP Statistics and Graphics Guide*.
- **Quantiles** and **Moments** outlines (for details, see "Continuous Variables," p. 85).

Bar Charts

Figure 3.6 Example of a Bar Chart



You can use a bar chart to visualize the distribution of a categorical variable. A bar chart looks similar to a histogram, since they both have bars that correspond to the levels of a variable. However, unlike a histogram, a bar chart shows a bar for every level of the variable.

Scenario

To use a bar chart, look at sample data for a group of companies. The information includes financial data for each company. For this example, focus on the type and size of the company.

Using a bar chart, you want to answer these questions:

- What is the most common type of company?
- What is the most common size of company?

Creating the Bar Chart

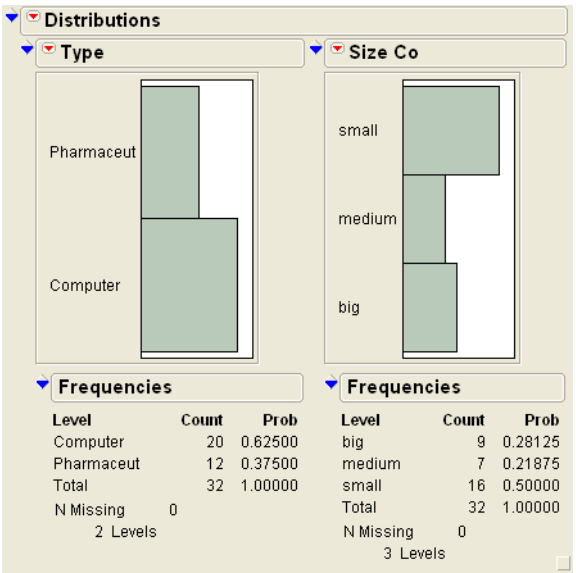
To answer your questions, create bar charts of Type and Size Co:

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Distribution**.
3 Assign Type and Size Co to the **Y, Columns** role.
4 Click **OK**. Figure 3.7 shows the bar charts.

Figure 3.7 Bar Charts of Type and Size Co

Bar Charts

Summary Information



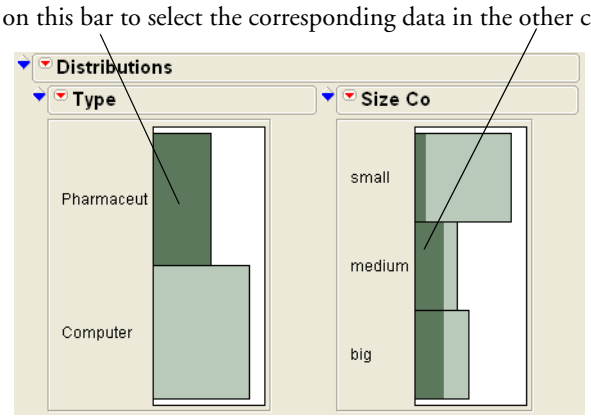
Interpreting the Bar Charts

From the bar chart you can see that there are more computer than pharmaceutical companies, and that the most common company size is small. The additional summary output gives detailed frequencies.

Interacting with the Bar Charts

As is the case with histograms, you can click on individual bars to highlight rows of the data table. Also, if more than one graph is created (as in Figure 3.7), clicking on a bar in one bar chart highlights the corresponding bar or bars in the other bar chart. For example, suppose you want to visualize the distribution of company size for the pharmaceutical companies. Click on the pharmaceutical bar in the Type bar chart, and the pharmaceutical companies are highlighted on the Size Co bar chart. See Figure 3.8. Also note that the corresponding rows in the data table are selected.

Figure 3.8 Clicking Bars

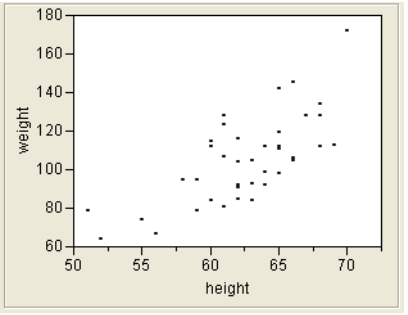


Using Multiple Variable Graphs

You can use multiple variable graphs to visualize the relationships and patterns between two or more variables. The graphs covered in this section include the following:

Scatterplots

Figure 3.9 Example of a Scatterplot



The scatterplot is the simplest of all the multiple variable graphs. You can use it to assess the relationship between two continuous variables. It is always the first step in assessing whether a correlation exists between two continuous variables. This section uses the Companies.jmp data.

Scenario

To use a scatterplot, look at sample data for a group of companies. The information includes financial data for each company. For this example, focus on sales and number of employees.

Using a bar chart, you want to answer these questions:

- What is the relationship between sales and the number of employees?
- Does sales increase with the number of employees?
- Can you predict average sales from the number of employees?

Creating the Scatterplot

To answer your questions, create a scatterplot of Sales ($M) versus # Employ by doing the following:

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Fit Y by X**.
3 Assign Sales ($M) to the **Y, Response** role.
4 Assign # Employ to the **X, Factor** role. Figure 3.10 shows the completed window.
5 Click **OK**. Figure 3.11 shows the scatterplot.

Figure 3.10 Fit Y by X Window

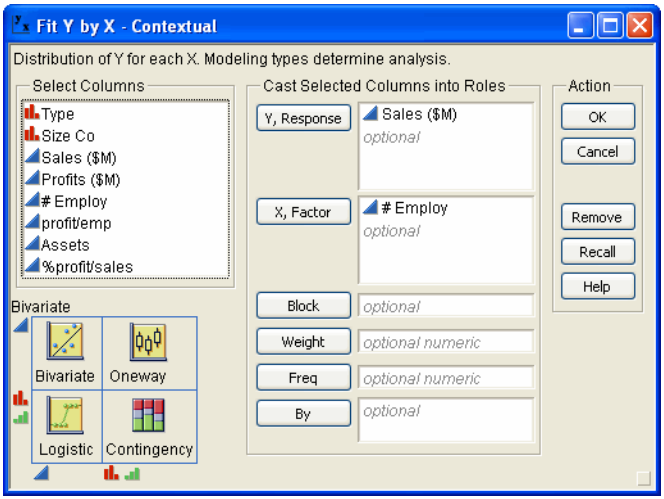
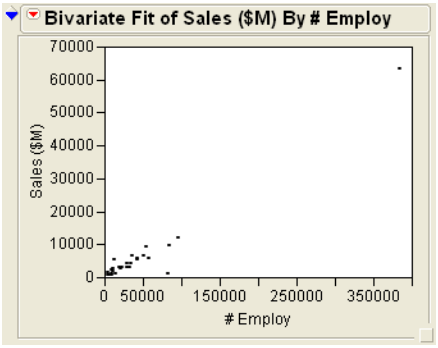


Figure 3.11 Scatterplot of Sales ($M) versus # Employ



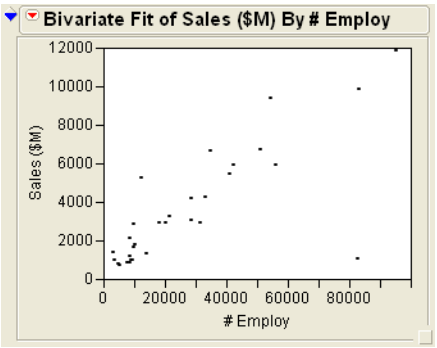
Interpreting the Scatterplot

There is one company that has a large number of employees and high sales. The data point representing this company makes it difficult to visualize the relationship between the variables. Remove the point from the plot and recreate the plot by taking the following steps:

1 Click on the point to select it.
2 Select **Rows** > **Exclude/Unexclude**. The data point is no longer included in calculations.
3 Select **Rows** > **Hide/Unhide**. The data point is hidden on all graphs.
4 On the bivariate red triangle menu, select **Script** > **Redo Analysis**.

The plot is recreated without the outlier. See Figure 3.12.

Figure 3.12 Outlier Removed



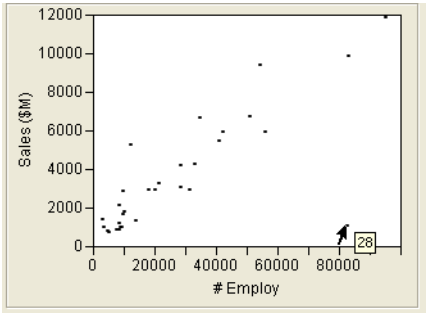
The scatterplot provides these answers:

- There is a relationship between the sales and the number of employees.
- Sales do increase with the number of employees, and the relationship is linear.
- You can predict average sales from the number of employees.

Interacting with the Scatterplot

As with other JMP graphics, the scatterplot is interactive. Hover over the point in the bottom right corner with the mouse to reveal the row number (in this example, 28). See Figure 3.13.

Figure 3.13 Hover Over a Point

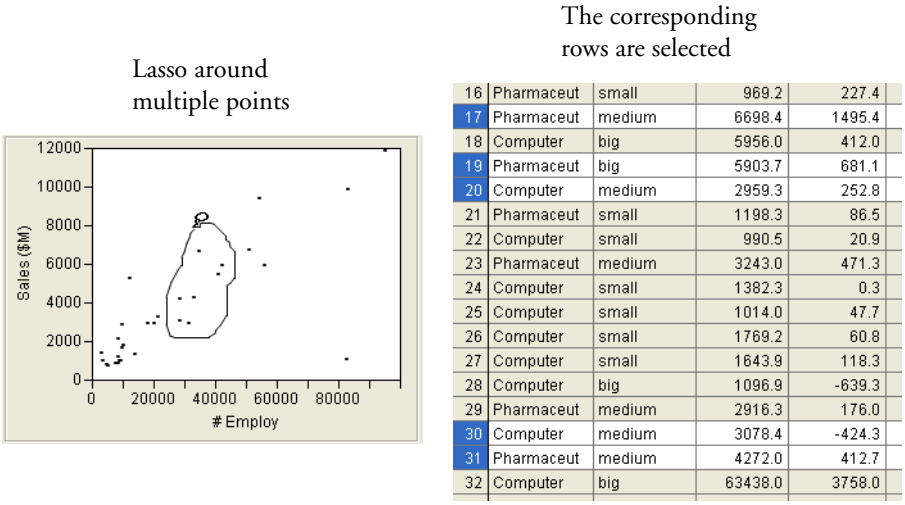


Click on a point to highlight the corresponding row in the data table. To select multiple points, do one of the following:

- Click and drag with the mouse around the points. This method enables you to select a rectangular region only.
- Use the lasso tool, shown in Figure 3.14. Select the lasso tool, and then click and drag around multiple points. See Figure 3.15. The lasso tool enables you to select an irregular-shaped region.

Figure 3.14 Lasso Tool

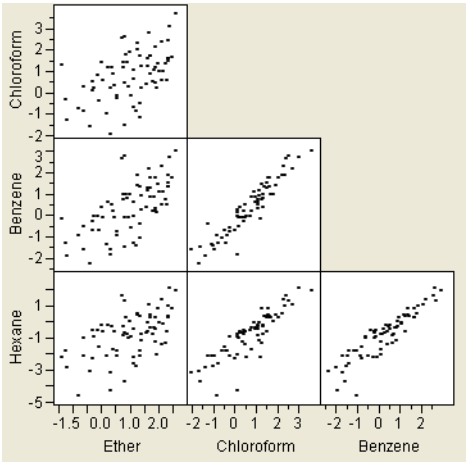
Figure 3.15 Select Points to Select Rows



16	Pharmaceut	small	969.2	227.4
17	Pharmaceut	medium	6698.4	1495.4
18	Computer	big	5956.0	412.0
19	Pharmaceut	big	5903.7	681.1
20	Computer	medium	2959.3	252.8
21	Pharmaceut	small	1198.3	86.5
22	Computer	small	990.5	20.9
23	Pharmaceut	medium	3243.0	471.3
24	Computer	small	1382.3	0.3
25	Computer	small	1014.0	47.7
26	Computer	small	1769.2	60.8
27	Computer	small	1643.9	118.3
28	Computer	big	1096.9	-639.3
29	Pharmaceut	medium	2916.3	176.0
30	Computer	medium	3078.4	-424.3
31	Pharmaceut	medium	4272.0	412.7
32	Computer	big	63438.0	3758.0

Scatterplot Matrix

Figure 3.16 Example of a Scatterplot Matrix



A scatterplot matrix is a collection of scatterplots organized into a grid (or matrix). Each scatterplot shows the relationship between a pair of variables. This plot enables you to quickly compare multiple variables at once.

Scenario

To show the use of a scatterplot matrix, look at data for chemical compounds. The chemical compounds were measured for solubility in different solvents. Using a scatterplot matrix, you want to answer these questions:

- Is there a relationship between any pair of variables?
- Which variables have the strongest relationship?

Creating the Scatterplot Matrix

To answer your questions, create a scatterplot matrix using Ether, Chloroform, Benzene, and Hexane:

1 Open the Solubility.jmp sample data table.
2 Select **Graph** > **Scatterplot Matrix**.
3 Assign Ether, Chloroform, Benzene, and Hexane to the **Y, Columns** role. Figure 3.17 shows the completed window.
4 Click **OK**. Figure 3.18 shows the scatterplot matrix.

Figure 3.17 Scatterplot Matrix Window

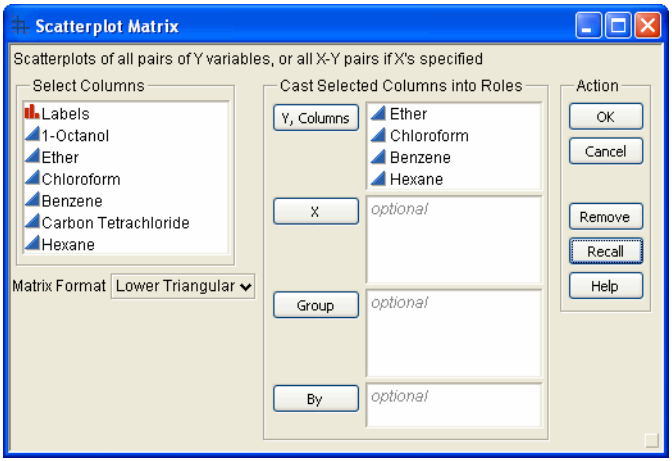
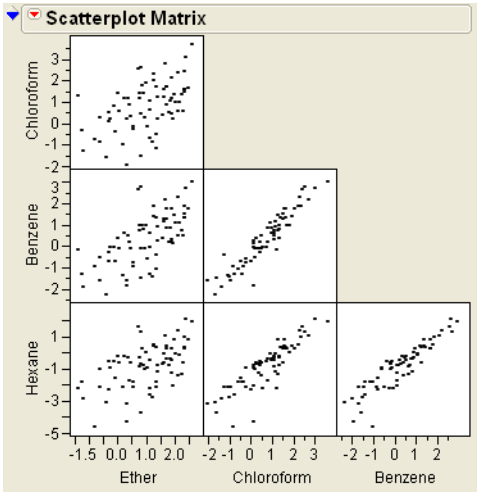


Figure 3.18 Scatterplot Matrix



Interpreting the Scatterplot Matrix

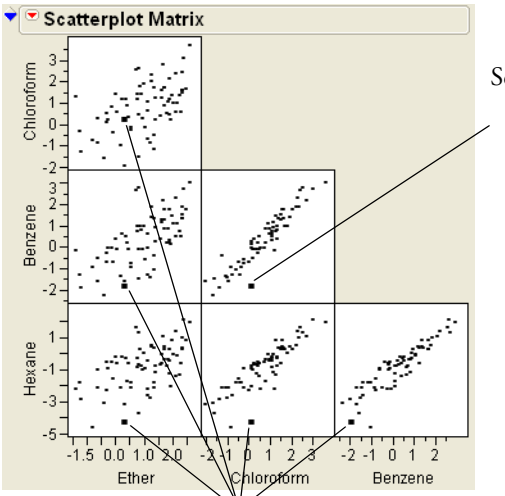
The scatterplot matrix shows these answers:

- All six pairs of variables are positively correlated. That is, as one variable increases, the other variable increases too.
- The strongest relationship appears to be between Benzene and Chloroform.

Interacting with the Scatterplot Matrix

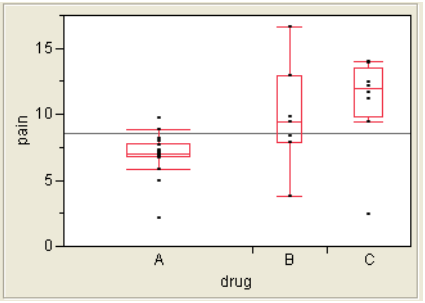
If you select a point in one scatterplot, it gets selected in all other scatterplots. For example, if you select a point in the Benzene versus Chloroform scatterplot, the same point gets highlighted in the other five plots. See Figure 3.19.

Figure 3.19 Highlighted Points



Side-by-Side Box Plots

Figure 3.20 Example of Side-by-Side Box Plots



Side-by-side box plots show you the relationship between one continuous variable and one categorical variable. You can assess differences in the continuous variable across levels of the categorical variable.

Scenario

To show the use of side-by-side box plots, use data consisting of pain measurements taken on patients using three different drugs. You want to answer the following questions:

- Are there differences in the average pain between the drugs?
- Are there differences in the variability in pain between the drugs?

Creating the Side-by-Side Box Plots

To answer your questions, create side-by-side box plots:

1 Open the Analgesics.jmp data table.
2 Select **Analyze** > **Fit Y by X**.
3 Assign pain to the **Y, Response** role.
4 Assign drug to the **X, Factor** role. Figure 3.21 shows the completed window.
5 Click **OK**.
6 Click on the red triangle, and select **Display Options** > **Box Plots**. Figure 3.22 shows the side-by-side box plots.

Figure 3.21 Fit Y by X Window

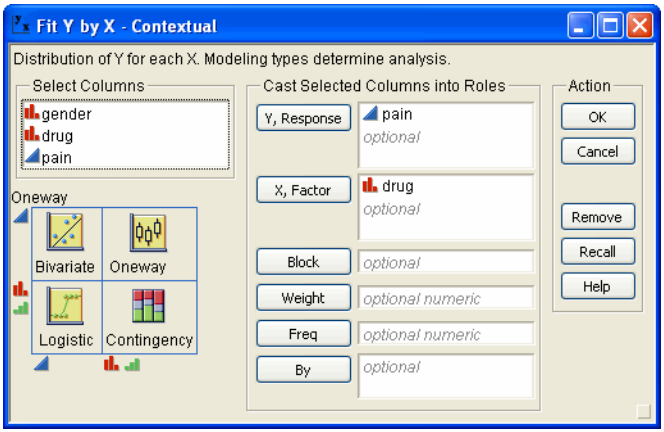
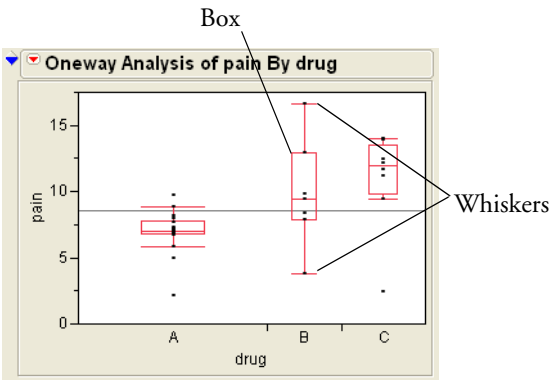


Figure 3.22 Side-by-Side Box Plots



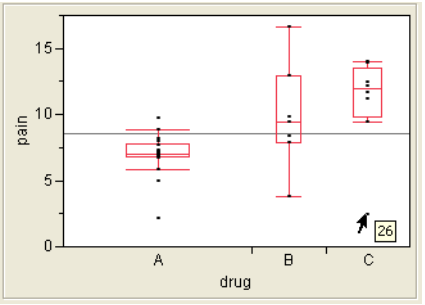
Interpreting the Side-by-Side Box Plots

The box plots are designed according to the following principles:

- The line through the box represents the median.
- The middle half of the data is within the boxes.
- The majority of the data falls between the ends of the whiskers.
- A data point outside the whiskers might be an outlier.
- Drug B appears to have higher variability than Drugs A and C.

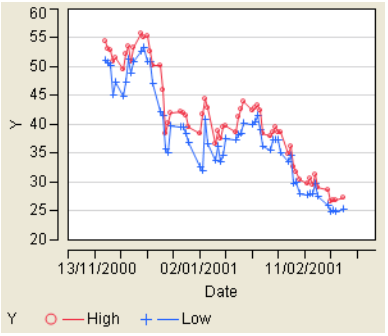
There is evidence to believe that patients on drug A feel less pain, since the box plot for drug A is lower than the others. Drug B appears to have higher variability than Drugs A and C, since the box plot is taller. There is one point for drug C that is a lot lower than the other drug C points. Hover over it with your mouse (see Figure 3.23) to see that it is row 26 of the data table. That point looks like it is more similar to the data in drug group A or B. The information in row 26 deserves investigation. There might have been a typographical error when the data was recorded.

Figure 3.23 Hovering Over the Outlier



Overlay Plots

Figure 3.24 Example of an Overlay Plot



Like scatterplots, overlay plots show the relationship between two or more variables. However, if one of the variables is time, the overlay plot does a better job at showing any trends in your data.

Note: To plot data over time, you can also use Graph Builder, bubble plots, control charts, and variability charts. For complete details, see the *JMP Statistics and Graphics Guide*.

Scenario

To show the use of an overlay plot, look at stock price data for a three month period. Using the overlay plot, you want to answer the following questions:

- Has the closing stock price changed over the last three months?
- How do the high and low stock prices relate to each other?

Creating the Overlay Plot

To answer your questions, create the overlay plots:

1 Open the Stock Prices.jmp data table.
2 Select **Graph** > **Overlay Plot**.

3 Assign Close to the **Y** role.

4 Assign Date to the **X** role. Figure 3.25 shows the completed window.

5 Click **OK**. Figure 3.26 shows the overlay plot.

Figure 3.25 Overlay Plot Window

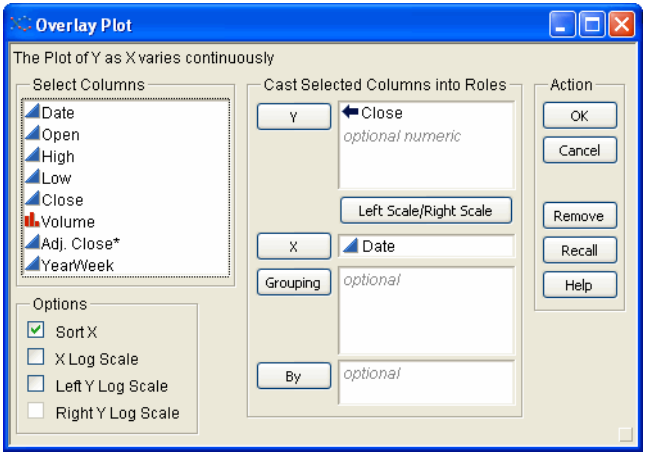
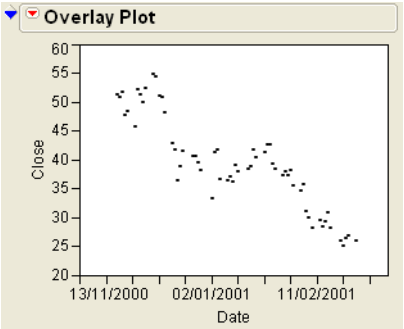


Figure 3.26 Overlay Plot of Close



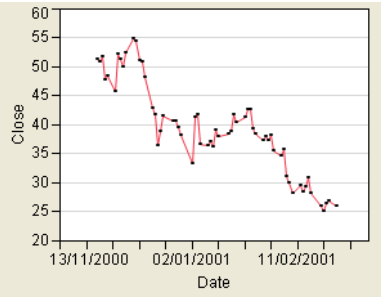
Interpreting and Interacting with the Overlay Plot

The overlay plot indicates that the closing stock price has been decreasing over the last several months. To see the trend more clearly, you can take the following actions:

- Connect the points with a line
- Add grid lines

To connect the points with a line, click the red triangle, and select **Connect Thru Missing**. To add grid lines, double-click on the Y axis, and check the box for Major Gridlines. Figure 3.27 shows the connected points and grid lines.

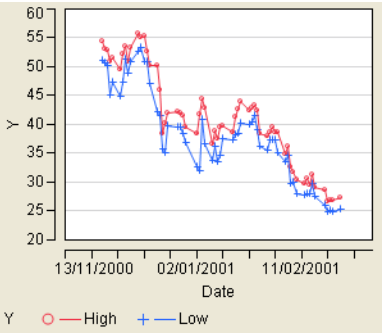
Figure 3.27 Connected Points and Grid Lines



Plotting Multiple Y Variables

More than one Y variable can be plotted on an overlay plot. For example, suppose you want to see both high and low price on the same plot. Follow the steps in "Creating the Overlay Plot," p. 64, this time assigning both High and Low to the **Y** role. Connect the points and add grid lines. Figure 3.28 shows the completed overlay plot.

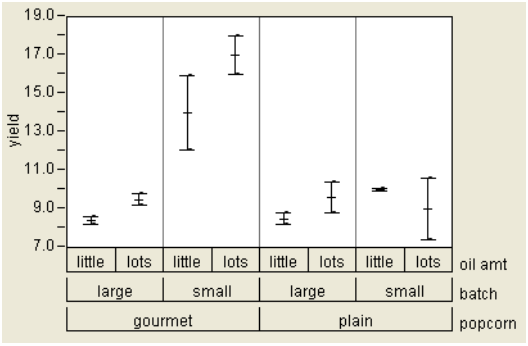
Figure 3.28 Two Y Variables



The legend at the bottom of the plot distinguishes the High and Low variables. From the overlay plot, you can see that the High price and Low price track each other very well.

Variability Chart

Figure 3.29 Example of a Variability Chart



In the graphs described so far, you can specify only a single X variable. Using a variability chart, you can specify multiple X variables. Using multiple X variables enables you to assess differences in means and variability across all the variables at once.

Scenario

To use a variability chart, look at sample data from a popcorn maker. Yield (volume of popcorn for a given measure of kernels) was measured for each combination of popcorn style, batch size, and oil amount. The popcorn maker is interested in the following question: Which combination of factors results in the highest popcorn yield?

Creating the Variability Chart

To answer the question, do the following:

1 Open the Popcorn.jmp data table.
2 Select **Graph** > **Variability/Gauge Chart**.
3 Assign yield to the **Y, Response** role.
4 Assign popcorn, batch, and oil amt, in that order, to the **X, Grouping** role. The order you assign the variables to the role is important, as it determines the order they appear on the variability chart. Figure 3.30 shows the completed window.
5 Click **OK**. Figure 3.31 shows the variability chart.

Figure 3.30 Variability Chart Window

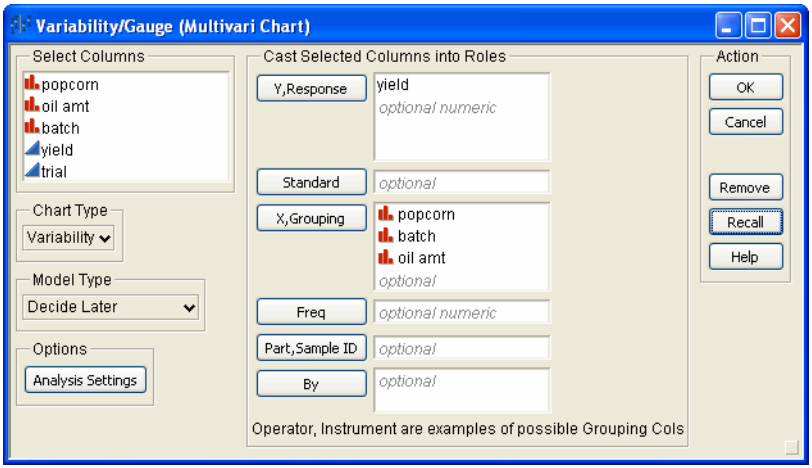
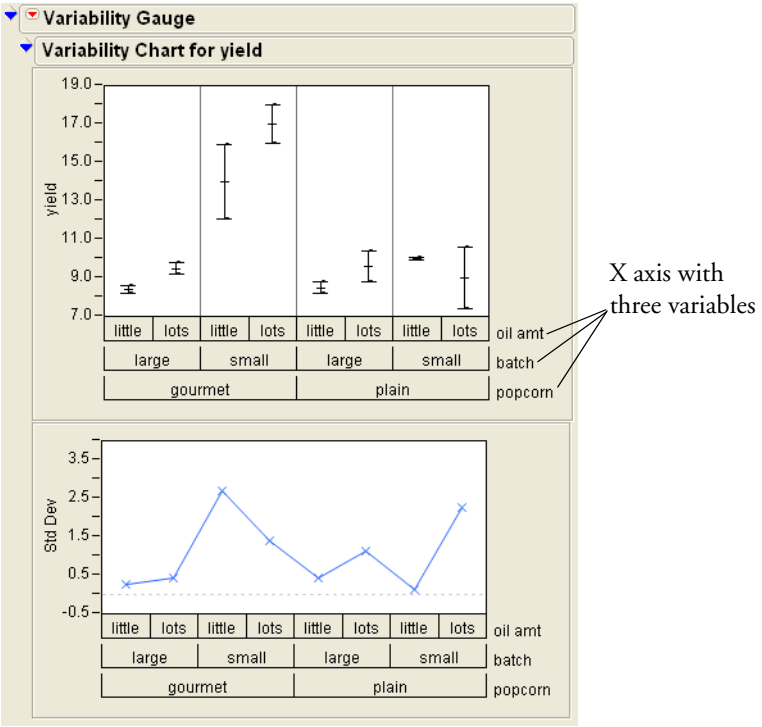


Figure 3.31 Results Window



Interpreting the Variability Chart

The top chart is the variability chart, and shows the yield broken down by each combination of the three variables. The bottom chart shows the standard deviation for each combination of the three variables. To hide this chart, click on the red triangle and deselect **Std Dev Chart**.

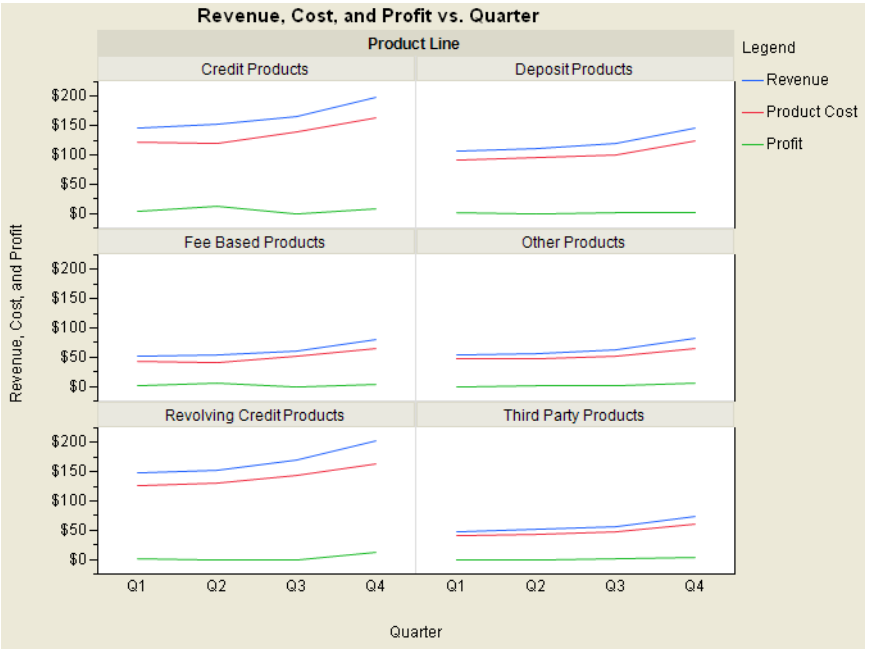
To answer the popcorn maker's question, look at the top chart showing yield. This chart shows that small gourmet batches produce the highest yield. To be more specific, you might ask the question: Is the yield high because those batches are small, or because those batches are gourmet? Note the following also learned from the chart:

- The yield from small, plain batches is low.
- The yield from large, gourmet batches is low.

Given this, it appears that only the combination of small and gourmet at the same time results in batches with high yield. This conclusion would have been impossible to reach with a chart that allowed only a single variable.

Graph Builder

Figure 3.32 Example of a Graph that was Created with Graph Builder



Using Graph Builder you can interactively create and modify graphs. To create the graphs that have been discussed so far, you launch a platform and specify variables. After the graph is created, you cannot change the variables. If you want to create a different type of graph, you have to launch another platform.

Use Graph Builder to accomplish the following tasks:

- Change variables by dragging and dropping them in and out.
- Create a different type of graph with a few mouse clicks.
- Partition the graph horizontally or vertically.

Note: Only some of the Graph Builder features are covered here. For complete details, see the *JMP Statistics and Graphics Guide*.

Scenario

To use Graph Builder, look at sample data of profit for multiple products. Suppose you are a business analyst and need to better understand the profitability of your product lines. You want to create a line plot that displays Revenue, Product Cost, and Profit, broken out by Product Line.

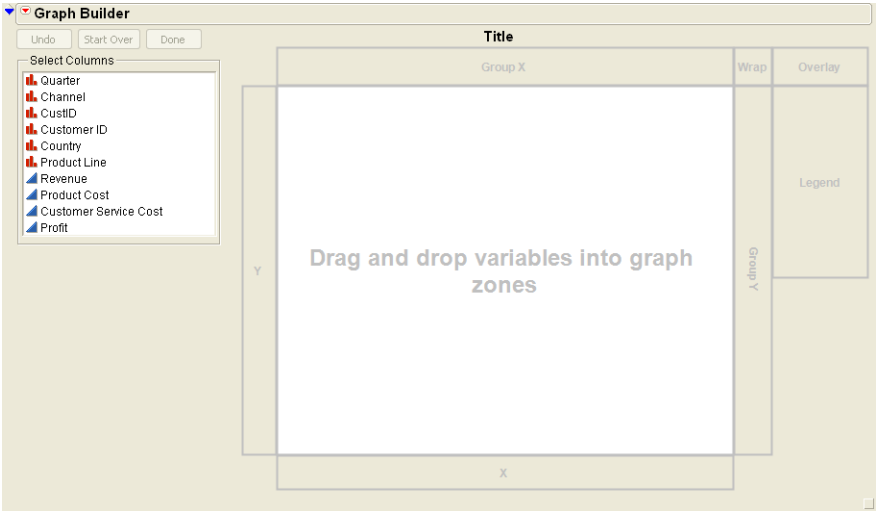
Creating the Graph

To better understand profitability, create the graph as follows:

1 Open the Profit by Product.jmp data table.

2 Select **Graph** > **Graph Builder**. The Graph Builder workspace appears. See Figure 3.33.

Figure 3.33 Graph Builder Workspace



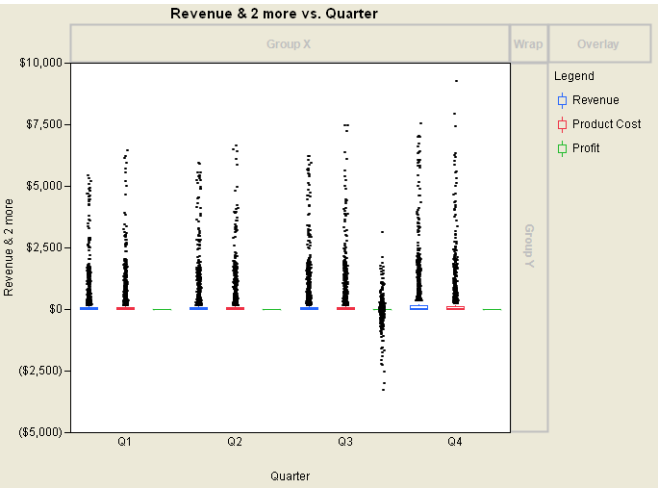
3 Select Quarter in the Select Columns box.

4 Click anywhere in the **X** zone to assign Quarter as the *X* variable.

5 Select Revenue, Product Cost, and Profit in the **Select Columns** box.

6 Click anywhere in the **Y** zone to assign all three variables as *Y* variables. See Figure 3.34.

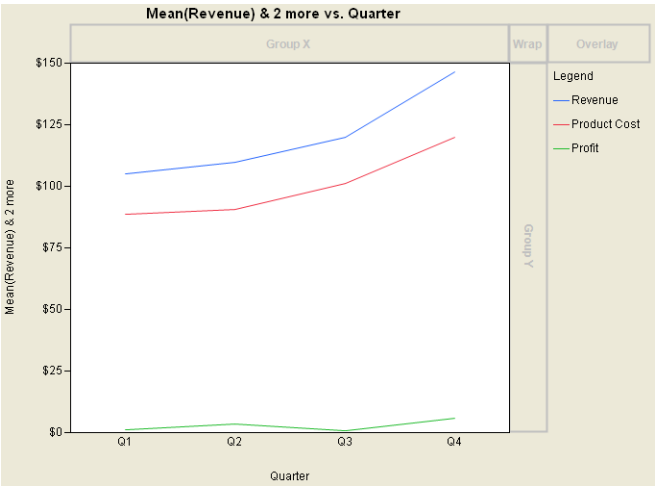
Figure 3.34 After Adding Y and X Variables



For the variables you are using, JMP chooses side-by-side box plots. Change the box plots to a line plot:

7 Right-click on the plot and select **Box Plot** > **Change To** > **Line**. The chart type changes to a line plot, as shown in Figure 3.35.

Figure 3.35 Line Plot

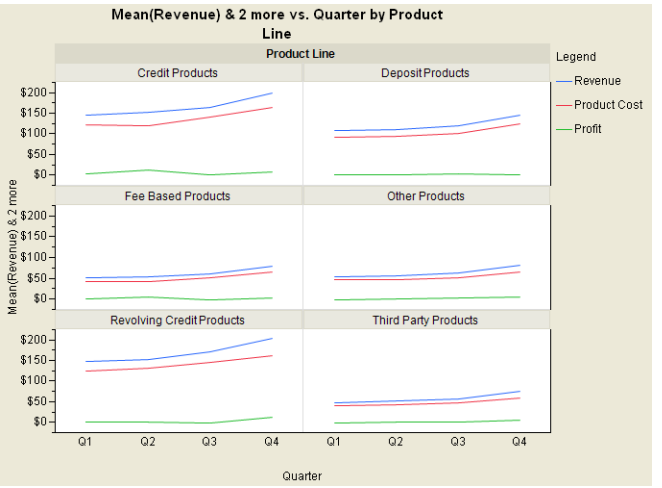


Now, create a separate chart for each product:

8 Select Product Line in the **Select Columns** box.

9 Click the **Wrap** zone to assign Product Line to it. A separate line plot is created for each product, as shown in Figure 3.36.

Figure 3.36 Final Line Plot

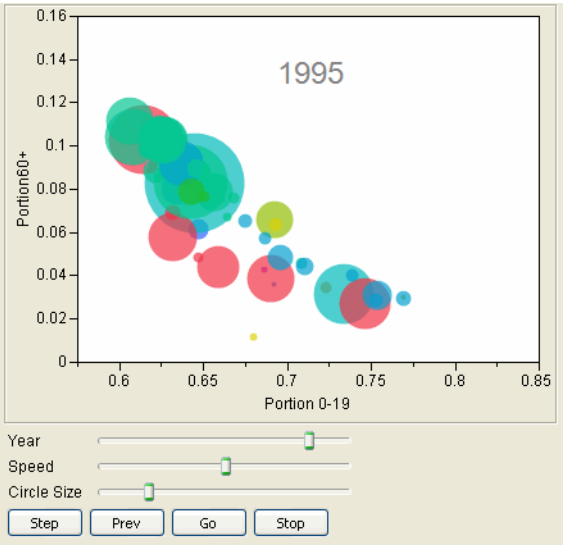


Interpreting the Graph

The plots show revenue, cost, and profit, broken down by product line. You can clearly see that some product lines produce more revenue than others.

Bubble Plots

Figure 3.37 Example of a Bubble Plot



A bubble plot is a scatterplot that represents its points as bubbles (or circles). You can change the size and color of the bubbles, and even animate them over time. With the ability to represent up to five

dimensions (*x* position, *y* position, size, color, and time), a bubble plot can produce dramatic visualizations and make data exploration easy.

Scenario

To use a bubble plot, look at population statistics. The statistics are for 116 countries or territories between the years 1950 to 2004. Total population numbers are broken out by age group, and not every country has data for every year. Using a bubble plot, you want to compare the portion of the population that is 19 years or younger to the portion that are 60 years or older. You want to determine how the relationship changes over time.

Creating the Bubble Plot

To answer your question, create the bubble plot:

1 Open the PopAgeGroup.jmp data table.
2 Select **Graph** > **Bubble Plot**.
3 Assign Portion60+ to the **Y** role. This corresponds to the *Y* variable on the bubble plot.
4 Assign Portion 0-19 to the **X** role. This corresponds to the *X* variable on the bubble plot.
5 Assign Country to the **ID** role. This tells JMP how to aggregate the rows of the data table into bubbles. Each unique level of the ID variable is represented by a bubble on the plot.
6 Assign Year to the **Time** role. This controls the time indexing when the bubble plot is animated.
7 Assign Pop to the **Sizes** role. This controls the size of the bubbles.
8 Assign Region to the **Coloring** role. This controls the colors of the bubbles. Figure 3.38 shows the completed window.
9 Click **OK**. Figure 3.39 shows the initial bubble plot.

Figure 3.38 Bubble Plot Window

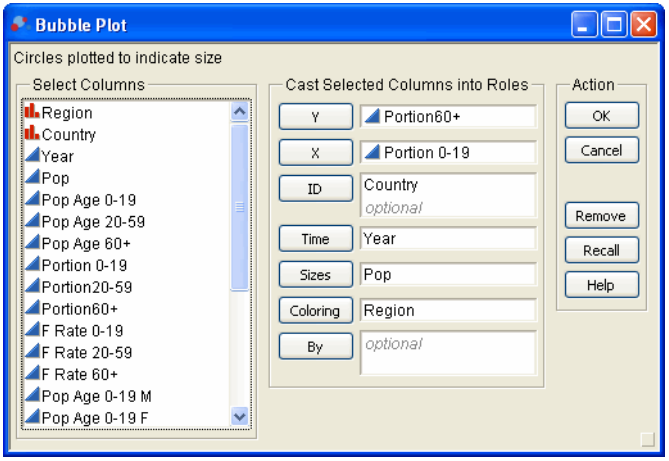
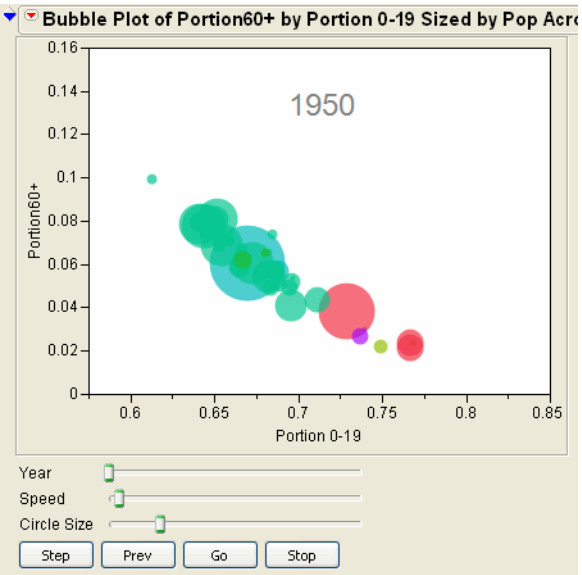


Figure 3.39 Initial Bubble Plot



Interpreting the Bubble Plot

Because the time variable (in this case, year) starts in 1950, the initial bubble plot shows the data for 1950. You can animate the bubble plot to cycle through all the years. Each successive bubble plot shows the data for that year. The data for each year determines the following:

- The *X* and *Y* coordinates
- The bubble's sizes
- The bubble's coloring
- Bubble aggregation

Note: For detailed information about how the bubble plot aggregates information across multiple rows, see the *JMP Statistics and Graphics Guide*.

The bubble plot for 1950 shows that if a country's population has a high proportion of people age 19 or younger, then that country generally has a low proportion of age 60 or older.

Click **Go** to animate the bubble plot through the range of years. Note that as time progresses, the Portion 0-19 decreases and the Portion60+ increases.

Stop stops the animation.

Step manually control the animation forward one unit of time.

Prev manually controls the animation back one unit of time.

Year is used to change the time index manually.

Speed controls the speed of the animation.

Circle Size controls the absolute sizes of the bubbles, while maintaining the relative sizes.

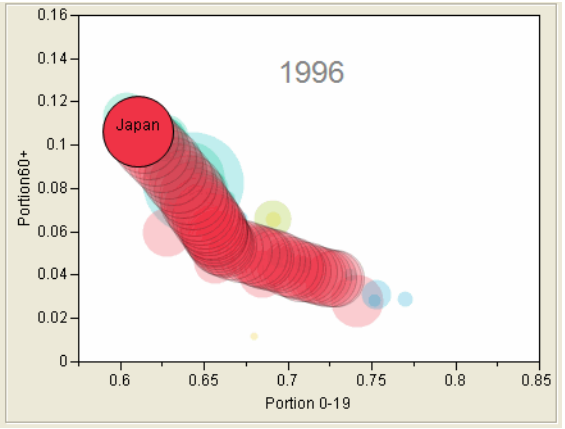
Interacting with the Bubble Plot

You can select a bubble to see the trend for that bubble over time. For example, in the 1950 plot, the large red bubble in the middle is Japan. To see the pattern of Japan through the years:

1 Click in the middle of the Japan bubble to select it.
2 Click the red triangle and select **Trail Bubbles** (the **Trail Lines** option shows the history as a line).
3 Click **Go**.

As the animation progresses through time, the Japan bubble leaves a trail of bubbles so you can see the history. See Figure 3.40.

Figure 3.40 Japan History



Focusing on the Japan bubble, you can easily see that over time, the proportion of the population 19 years old or less decreased, and the proportion of the population 60 years old or more increased.

Chapter 4

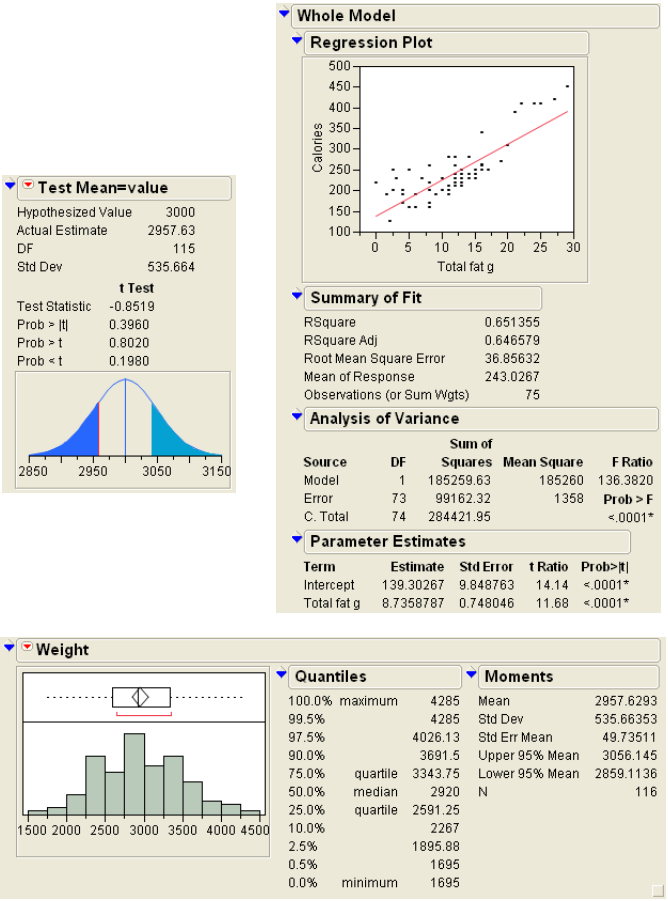
Analyzing Your Data

Distributions, Relationships, and Models

Analyzing your data helps you make informed decisions. Data analysis often involves these actions:

- Examining distributions
- Assessing relationships
- Hypothesis testing
- Building models

Figure 4.1 Analysis Examples



Contents

About This Chapter

Statistical and analytical methods are very important to any organization seeking to be data driven. This chapter contains the following information:

Prerequisites

Before you analyze your data, there are two things that you need to review:

The Importance of Graphing Your Data

Graphing your data (data visualization) is important to any data analysis, and should always occur before the use of statistical tests or model building. To illustrate why data visualization should be an early step in your data analysis process, consider the following example:

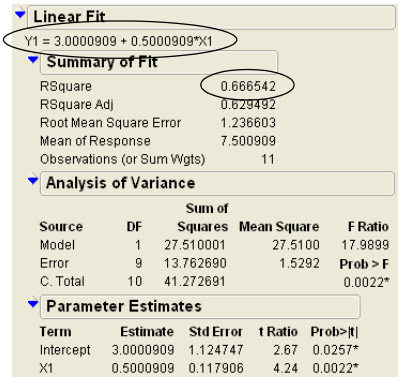
1 Open the Anscombe.jmp data table (F. J. Anscombe (1973), *American Statistician*, 27, 17-21). Note the data consists of four pairs of *X* and *Y* variables.

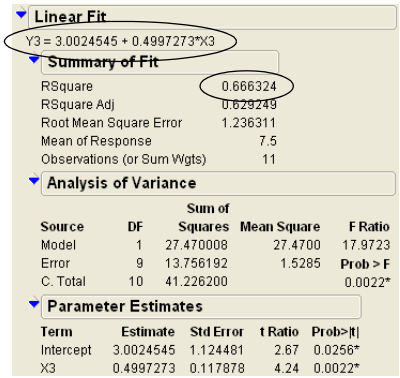
2 In the table panel, click on the red triangle next to The Quartet and select **Run Script**.

The script creates a simple linear regression on each pair of variables using **Fit Y by X**. The **Show Points** option is turned off, so that none of the data can be seen on the scatterplots. Figure 4.2 shows the model fit and other summary information for each regression.

Figure 4.2 Four Models

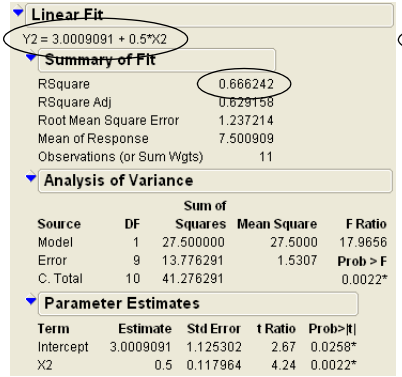
Model for pair 1

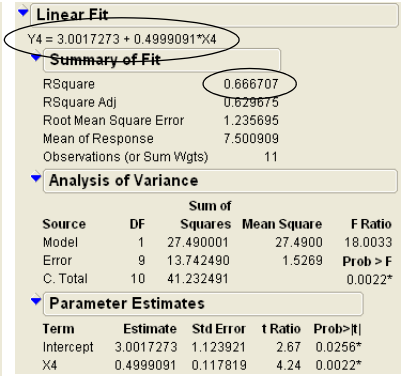




Model for pair 3

Model for pair 2





Model for pair 4

Notice that all four models and RSquare values are nearly identical. The fitted model in each case is essentially $Y = 3 + 0.5X$, and the RSquare value in each case is essentially 0.66. If your data analysis took into account only the above summary information, you would likely conclude that the relationship between X and Y is the same in each case. However, at this point, you have not visualized your data. Your conclusion might be wrong.

To visualize the data, add the points to all four scatterplots:

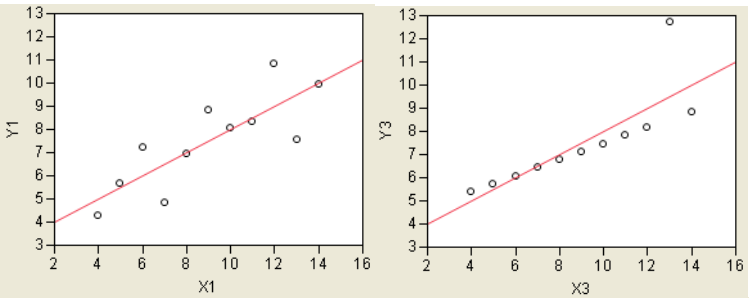
1 Hold down the Ctrl key.

2 On one of the four Bivariate Fit outlines, click the red triangle.

3 Select **Show Points**. Figure 4.3 shows the points added to the plots.

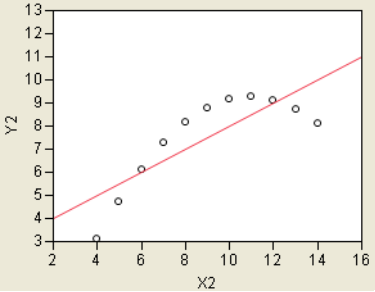
Figure 4.3 Scatterplots with Points Added

Plot for pair 1

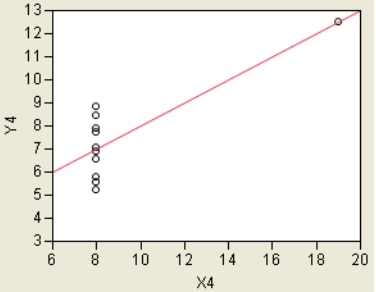


Plot for pair 3

Plot for pair 2



Plot for pair 4



The scatterplots show that the relationship between *X* and *Y* is not the same for the four pairs:

- Plot 1 represents a nice, linear relationship.
- Plot 2 represents a non-linear relationship.
- Plot 3 represents a linear relationship, except for one outlier.
- Plot 4 has all the data at x = 8, except for one point.

This example illustrates that conclusions that are based on statistics alone can be inadequate. A visual exploration of the data should be an early part of any data analysis.

Understanding Modeling Types

Data in JMP can be of different types. JMP refers to this as the modeling type of the data. There are three modeling types in JMP:

Table 4.1 Modeling Types

Modeling Type	Description	Examples	Specific Example
Continuous	Numeric only. Used in operations like sums and means.	Height Temperature Time	Time to complete a test might be 2 hours, or 2.13 hours.
Ordinal	Numeric or character. Values belong to ordered categories.	Month (1,2,...,12) Letter grade (A, B,...F) Size (small, medium, large)	Month of the year can be 2 (February) or 3 (March), but not 2.13.

Table 4.1 Modeling Types

Modeling Type	Description	Examples	Specific Example
Nominal	Numeric or character. Values belong to categories, but the order is not important.	Gender (M or F) Color Test result (pass or fail)	Gender can be M or F, with no order. Gender categories can also be represented by a number (M=1 and F=2).

Example of Modeling Type Results

Different modeling types produce different results in JMP. To see an example of the differences, follow these steps:

1 Open the Big Class.jmp sample data table.
2 Select **Analyze** > **Distribution**.
3 Assign height and age to the Y, Columns role.
4 Click **OK**. Figure 4.4 shows the results.

Figure 4.4 Distribution Results for height and age

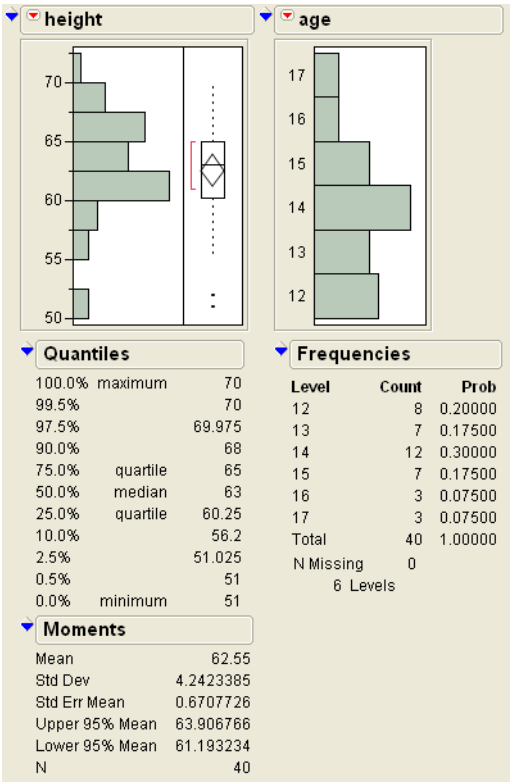


Table 4.2 compares the differences between the results for height and age.

Table 4.2 Results for height and age

Variable	Modeling Type	Results
height	Continuous	Histogram, Quantiles, and Moments
age	Ordinal	Bar chart and Frequencies

Changing the Modeling Type

If you want to treat a variable differently, you can change the modeling type. For example, in Figure 4.4, the modeling type for age is ordinal. If you want to find the average age instead of frequency counts, change the modeling type to continuous and create the distribution:

1 Double-click on the age column heading. The Column Info window appears.
2 Change the Modeling Type to **Continuous**. See Figure 4.5.
3 Click **OK**.
4 Repeat the steps in the example (see "Example of Modeling Type Results," p. 82) to create the distribution. Figure 4.6 shows the distribution results when age is ordinal and continuous.

Figure 4.5 Column Info Window

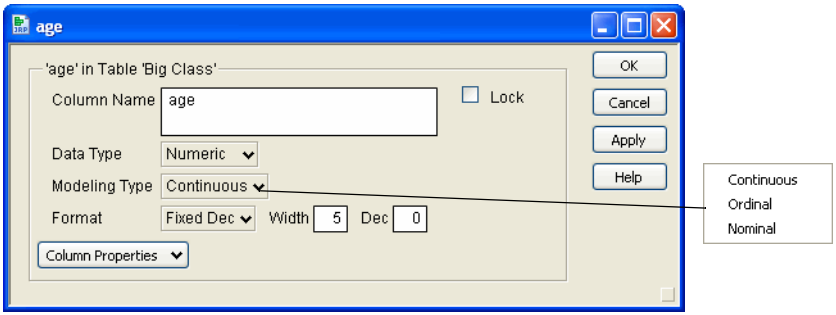
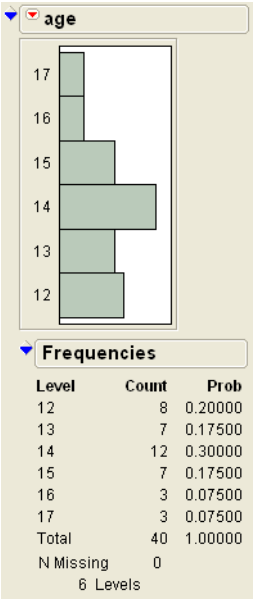
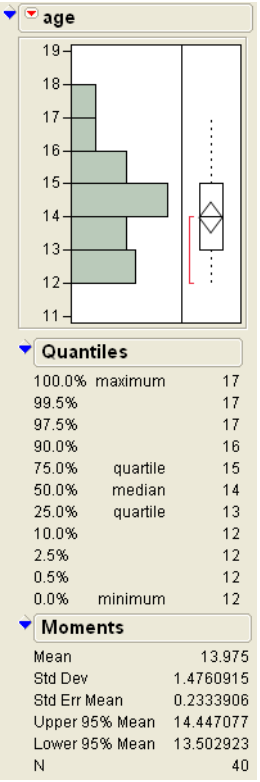


Figure 4.6 Different Modeling Types for age

Ordinal



Continuous



Performing Analyses

Performing statistical analyses normally involves either:

- "Analyzing Distributions," p. 84
- "Analyzing Relationships," p. 91

Analyzing Distributions

Analyzing a single variable usually involves examining the distribution of the variable. You can analyze distributions using the **Distribution** platform.

Note: For complete details of the **Distribution** platform, see the *JMP Statistics and Graphics Guide*.

This section examines distributions for the following:

- "Continuous Variables," p. 85
- "Categorical Variables," p. 89

Continuous Variables

Analyzing a continuous variable might include questions such as the following:

- What is the average of the data?
- Is the average statistically different from a target or historical value?
- How spread out are the data? In other words, what is the standard deviation?
- What are the minimum and maximum values?
- Are there any outliers in the data?
- Does the shape of the data match any known distributions?

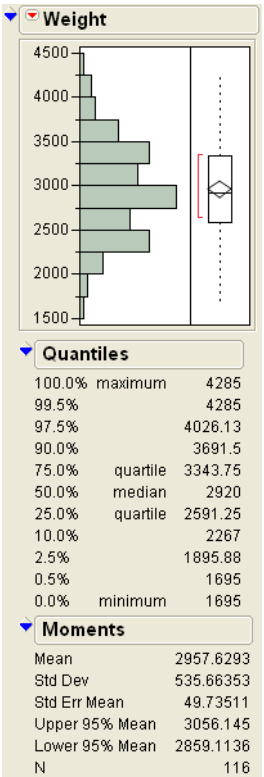
You can answer these and other questions with graphs, summary statistics, and simple statistical tests. You can analyze a continuous variable to see what you can learn about the data.

Example of Analyzing a Continuous Variable

Suppose you are an analyst for a railroad company. Your company plans to start transporting cars. You want to analyze the distribution of car weights:

1. Open the Car Physical Data.jmp sample data table. The data table contains information about 116 models of cars.
2. Select **Analyze** > **Distribution**.
3. Assign Weight to the **Y, Columns** role.
4. Click **OK**. Figure 4.7 shows the initial results.

Figure 4.7 Distribution of Weight

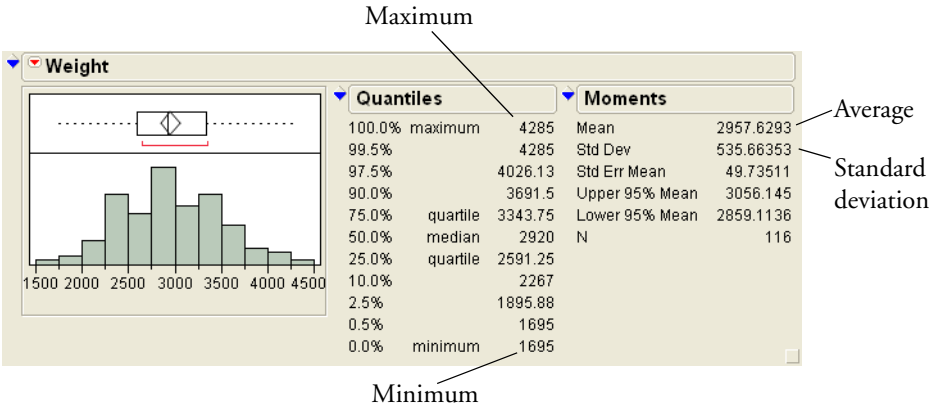


The default report window contains three sections:

- A histogram and a box plot to visualize the data.
- A Quantiles outline that shows the percentiles of the distribution.
- A Moments outline that shows the mean, standard deviation, and other statistics.

To rotate the report, click on the red triangle next to **Weight**, and select **Display Options** > **Horizontal Layout**. See Figure 4.8.

Figure 4.8 Horizontal Layout



Interpreting the Distribution Results

Using the results presented in Figure 4.8, you can answer some of the questions that were posed earlier:

Table 4.3 Questions Answered Using the Report Window Results*

Question	Histogram	Quantiles	Moments
What is the average car weight?	~3000 lbs		2957 lbs
How spread out are the weights (std dev)?			535 lbs
What are the minimum and maximum weights?	~1500 lbs and ~4500 lbs	1695 lbs and 4285 lbs	
Are there any outliers?	No		

*Blank = not applicable

The default report, shown in Figure 4.8, provides a minimal set of graphs and statistics. You can add many more graphs and statistics to the report window. Here are a few examples of actions you can perform:

- "Performing a t-test," p. 87
- "Fitting Distributions," p. 88

Performing a t-test

A t-test helps you use sample data to draw inferences about the broader population. For example, you can use the sample of 116 car models to draw conclusions about all car models. Suppose you want to test whether the average car weight in the population is equal to a target value. The target value is 3000

pounds, since the railroad company you work for has determined that an average weight of 3000 pounds is the most efficient.

Perform the t-test as follows:

1 Click on the red triangle next to **Weight**, and select **Test Mean**. The Test Mean window opens.
2 In the Specify Hypothesized Mean box, type 3000. See Figure 4.9.
3 Click **OK**. Figure 4.10 shows the Test Mean results.

Figure 4.9 Test Mean Window

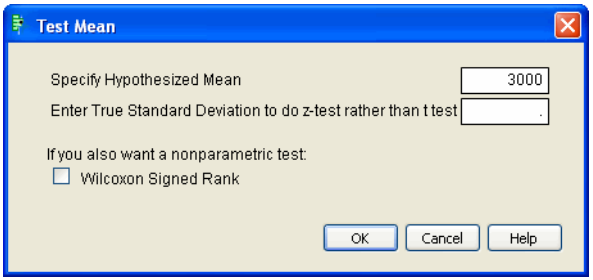
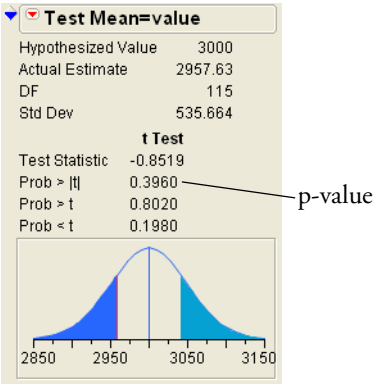


Figure 4.10 Test Mean Results

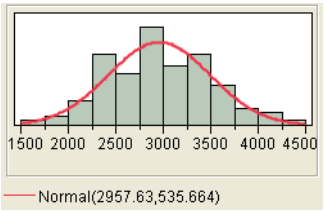


The primary result of the t-test is the p-value. In this example, the p-value is 0.396. Using a significance level of 0.05, and noting that 0.396 is greater than 0.05, you can conclude that the average weight of car models in the broader population is not significantly different from 3000 lbs.

Fitting Distributions

You can make predictions about a population using a fitted distribution. For this example, you can fit a distribution to estimate the proportion of cars in the broader population that weigh more than a certain amount. Fit a Normal distribution to the Weight variable. Click on the red triangle next to **Weight**, and select **Continuous Fit** > **Normal**. Figure 4.11 shows the fitted Normal curve overlaid on the histogram, and the Fitted Normal outline added to the report.

Figure 4.11 Normal Fit



Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	2957.6293	2859.1136	3056.145
Dispersion	σ	535.66353	474.47889	615.10733

-2log(Likelihood) = 1785.96718329299

Using this information, you can estimate the proportion of cars in the broader population that weigh more than 4000 lbs, the amount your railroad company considers to be heavy.

Categorical Variables

Analyzing a categorical variable might include questions such as the following:

- How many levels does the variable have?
- How many data points does each level have?
- What proportions of the total do each level represent?

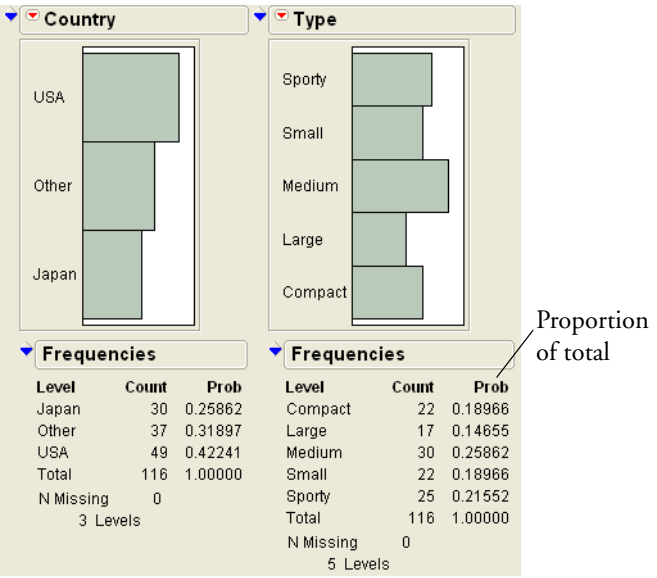
Note: Categorical variables have a modeling type of ordinal or nominal.

Example of Analyzing a Categorical Variable

Suppose you are an analyst for a railroad company. Your company plans to start transporting cars. You want to analyze the distribution of car type and country of origin:

1 Open the Car Physical Data.jmp sample data table. The data table contains information about 116 models of cars.
2 Select **Analyze** > **Distribution**.
3 Assign Country and Type to the **Y, Columns** role.
4 Click **OK**. Figure 4.12 shows the initial results.

Figure 4.12 Distribution for Country and Type



Interpreting the Distribution Results

The default report window contains two sections: a bar chart and a Frequencies outline. The bar chart is a graphical representation of the frequency information provided in the Frequencies outline. The Frequencies outline contains the following:

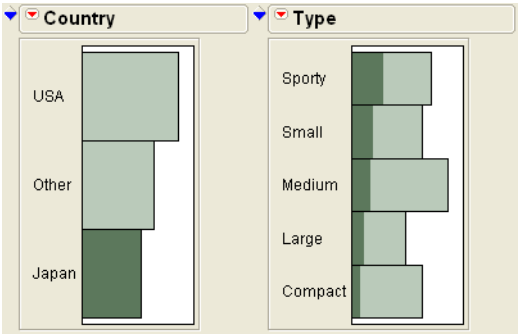
- Levels
- Total counts for each level
- Proportion of the total each level represents

For example, the Count for cars in the Compact level is 22, which accounts for about 19% of the 116 observations.

Interacting with the Distribution Results

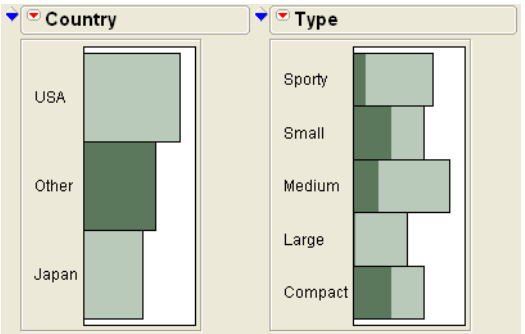
If you select a bar in one chart, the corresponding data in the other chart is highlighted. For example, select the Japan bar in the Country bar chart. In the Type bar chart, you can see that a large number of Japanese cars are sporty. See Figure 4.13.

Figure 4.13 Japanese Cars



If you select the Other category, you notice that a majority of cars are small or compact, and almost none are large. See Figure 4.14.

Figure 4.14 Other Cars



Analyzing Relationships

You can analyze the relationship between variables using either the **Fit Y by X** platform or the **Fit Model** platform.

Note: The basic platforms and options are covered here. For complete details and explanations of all platform options, see the *JMP Statistics and Graphics Guide*.

Table 4.4 shows the four primary types of relationships.

Table 4.4 Relationship Types

X	Y	Section
Continuous	Continuous	• "Using Regression with One Predictor," p. 92 • "Using Regression with Multiple Predictors," p. 102
Categorical	Continuous	• "Comparing Averages for One Variable," p. 97 • "Comparing Averages for Multiple Variables," p. 106
Categorical	Categorical	"Comparing Proportions," p. 100
Continuous	Categorical	This is an advanced topic. See the *JMP Statistics and Graphics Guide*.

Using Regression with One Predictor

You can use regression to analyze the relationship between two continuous variables. A regression model predicts the average value of one variable (*Y*) from the value of another variable (*X*). The *X* variable is also called a predictor.

Scenario

To use regression, look at financial data for 32 companies from the pharmaceutical and computer industries. Intuitively, it makes sense that companies that have more employees can generate more sales revenue than companies that have fewer employees. You can use the financial data to accomplish the following tasks:

- "Assess the Relationship," p. 92
- "Fit the Regression Model," p. 94
- "Predict Average Sales," p. 95

Assess the Relationship

To assess the relationship between the number of employees and the amount of sales revenue, create a scatterplot as follows:

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Fit Y by X**.
3 Assign Sales ($M) to the **Y, Response** role.
4 Assign # Employ to the **X, Factor** role. Figure 4.15 shows the completed window.
5 Click **OK**. Figure 4.16 shows the initial results.

Figure 4.15 Completed Fit Y by X Window

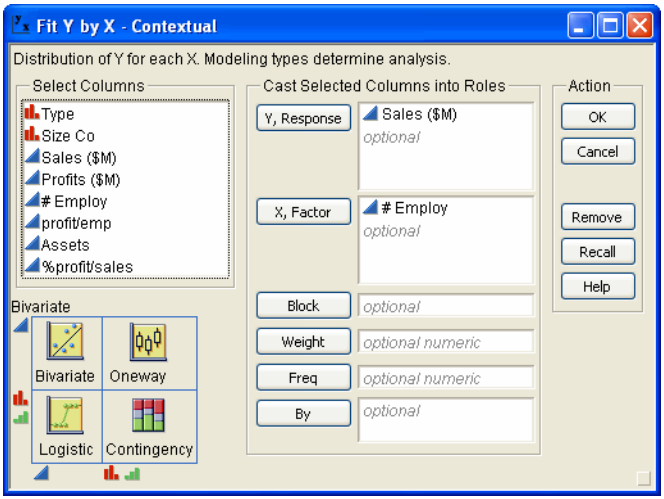
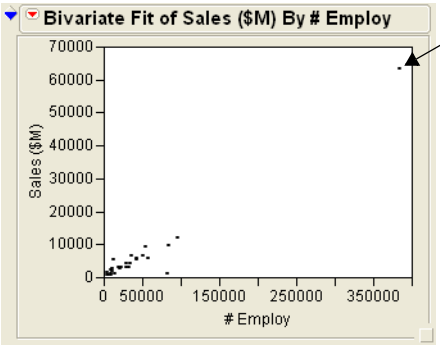


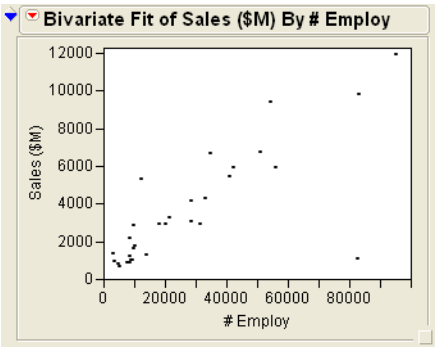
Figure 4.16 Scatterplot of Sales versus Employees



The scatterplot results show that there is one company with significantly more employees and higher sales than the other companies. The data point representing this company is an outlier. The outlier is stretching out the scale of the plot, making it difficult to examine the relationship. Exclude and hide that data point:

1 Click on the data point.
2 Select **Rows** > **Exclude/Unexclude**. The data point is no longer included in calculations.
3 Select **Rows** > **Hide/Unhide**. The data point is hidden on all graphs.
4 To recreate the scatterplot, click on the red triangle next to **Bivariate Fit**, and select **Script** > **Redo Analysis**. Figure 4.17 shows the updated scatterplot.

Figure 4.17 Updated Scatterplot

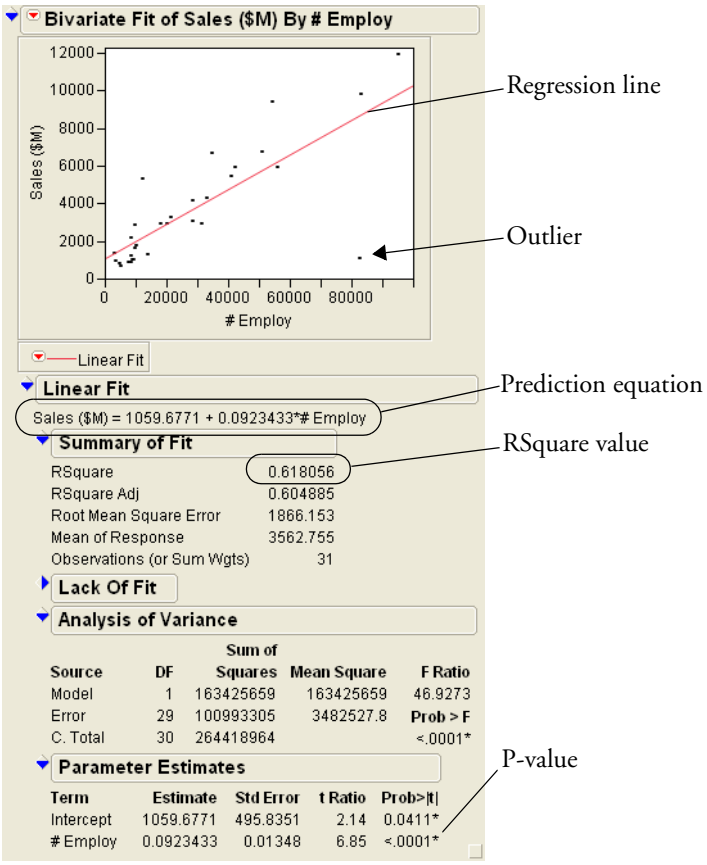


The updated scatterplot provides a clearer picture of the relationship between sales and the number of employees. As expected, the more employees a company has, the higher sales it can generate.

Fit the Regression Model

In order to predict the sales revenue from the number of employees, fit a regression model. Click on the red triangle next to **Bivariate Fit** and select **Fit Line**. The regression line is added to the scatterplot and outlines are added to the report. See Figure 4.18.

Figure 4.18 Regression Line



Within the outlines, look at the following results:

- The p-value of <.0001
- The RSquare value of 0.618

From these results, you can conclude the following:

- Since the p-value is less than a significance level of 0.05, this means that including the number of employees in the prediction model significantly improves the ability to predict average sales.
- Since the correlation coefficient is the square root of the RSquare value, in this example it is 0.786. A correlation coefficient of 0 indicates no relationship between the two variables, and a correlation coefficient of 1 (or -1) indicates a perfect linear relationship.

Predict Average Sales

You can use the regression model to predict the average sales a company might expect if they have a certain number of employees. The prediction equation for the model is as follows:

Average sales = 1059.67 + 0.092*employees

For example, in a company with 70,000 employees, here is the equation:

$7,499.67 = 1059.67 + 0.092*70,000

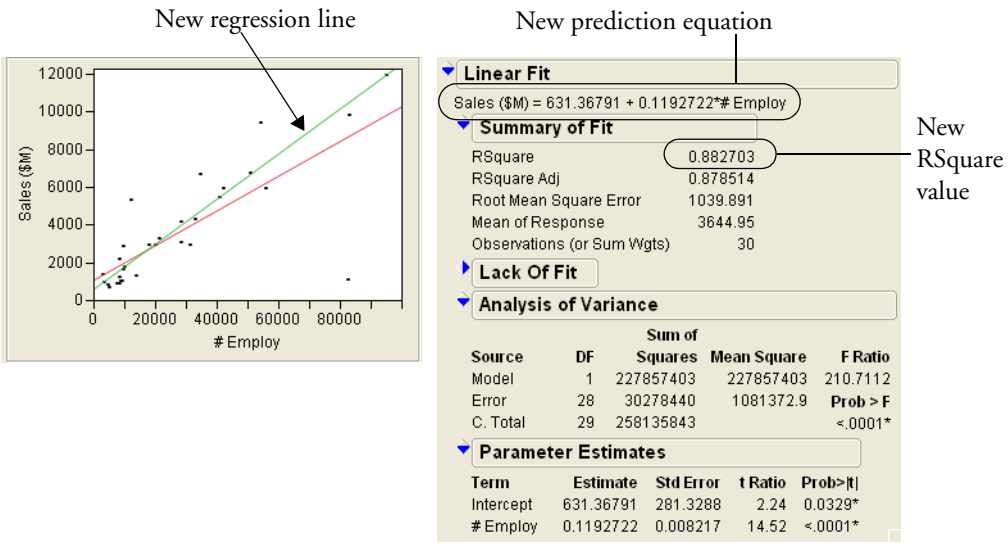
Look again at the scatterplot in Figure 4.18. There is an outlier that does not follow the general pattern of the other companies. You can find out if the prediction model changes if you exclude the outlier:

1 Click on the data point.
2 Select **Rows** > **Exclude/Unexclude**.
3 Re-fit the model. Click on the red triangle next to **Bivariate Fit** and select **Fit Line**.

The results are as follows. See Figure 4.19.

- A new regression line is added to the scatterplot.
- A new Linear Fit outline is added to the report, which includes:
  - A new prediction equation
  - A new RSquare value

Figure 4.19 Model without Outlier



Using the results in Figure 4.19, you can make the following conclusions:

- The outlier was pulling down the regression line for the larger companies, and pulling the line up for the smaller companies.
- The new model fits the data better, since the new correlation (0.94) is closer to 1 than the first correlation coefficient (0.786).

Using the new prediction equation, the predicted average sales for a company with 70,000 employees can be calculated as follows:

$8961.36 = 631.36 + 0.119*70,000

The total for the first model was $7499.67, so this is an increase of about $1460.

Comparing Averages for One Variable

If you have a continuous Y variable, and a categorical X variable, you might want to compare averages across levels of the X variable.

Scenario

To illustrate comparing averages, look at the financial data for 32 companies from the pharmaceutical and computer industries. Suppose you are interested in comparing the profits of computer companies to the profits of pharmaceutical companies. Use the financial data to:

- "Assess the Relationship," p. 97
- "Performing the t-test," p. 99

Assess the Relationship

To assess the relationship between profits and company type, create a plot:

1 Open the Companies.jmp sample data table.

Note: If you still have the Companies.jmp data file open from the previous example, you might have rows that are excluded or hidden. To return the rows to the default state (all rows included and none hidden), select **Rows** > **Clear Row States**.

---

2 Select **Analyze** > **Fit Y by X**.

3 Assign Profits ($M) to the **Y, Response** role.

4 Assign Type to the **X, Factor** role. Figure 4.20 shows the completed window.

5 Click **OK**. Figure 4.21 shows the initial results.

Figure 4.20 Completed Fit Y by X Window

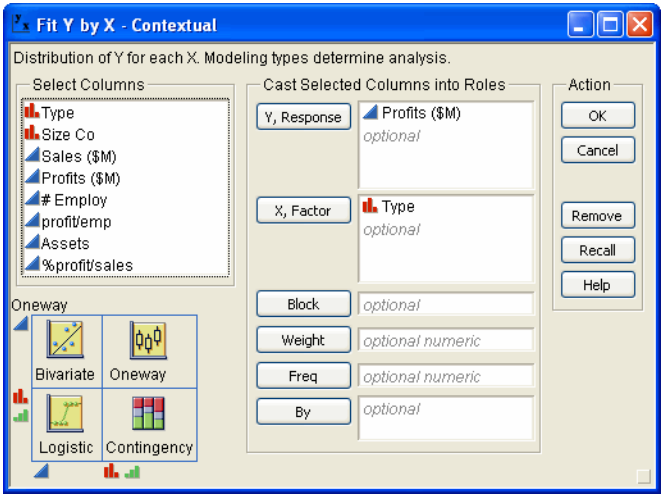
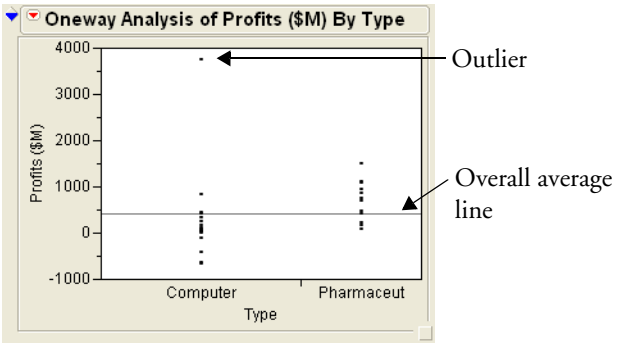


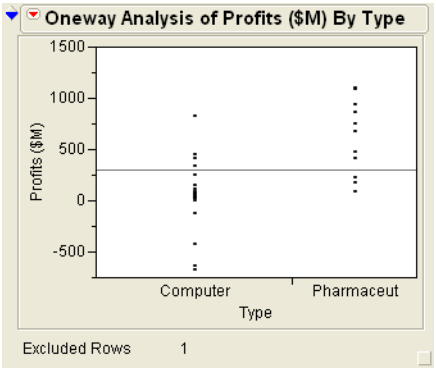
Figure 4.21 Profits by Company Type



There is an outlier in the Computer Type. The outlier is stretching the scale of the plot and making it difficult to compare the profits. Exclude and hide that data point:

1 Click on the data point.
2 Select **Rows** > **Exclude/Unexclude**. The data point is no longer included in calculations.
3 Select **Rows** > **Hide/Unhide**. The data point is hidden from all graphs.
4 To recreate the plot, click on the red triangle next to **Oneway Analysis**, and select **Script** > **Redo Analysis**. See Figure 4.22.

Figure 4.22 Updated Plot

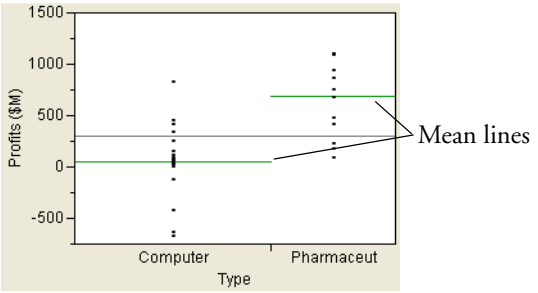


You can see that removing the outlier gives you a clearer picture of the data. To continue analyzing the relationship, click on the red triangle next to **Oneway Analysis** and select these options:

- **Display Options** > **Mean Lines**. This adds mean lines to the plot.
- **Means and Std Dev**. This displays a report that provides averages and standard deviations.

Figure 4.23 shows the mean lines and the report.

Figure 4.23 Mean Lines and Report



Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
Computer	19	55.753	357.909	82.11	-116.8	228.26
Pharmaceut	12	690.075	429.988	124.13	416.9	963.28

The updated plot shows that pharmaceutical companies have higher average profits than computer companies. In the report, if you subtract the mean values, the difference is about $635 million. Looking at the plot, you can also see that some of the computer companies have negative profits, while all of the pharmaceutical companies have positive profits.

Performing the t-test

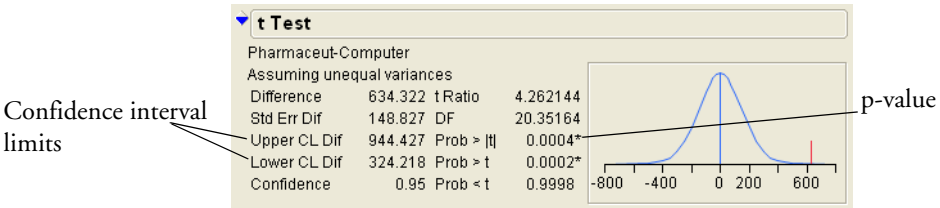
Because you have looked at only a sample of companies, you need to ask these questions:

- Does a difference exist in the larger population of all companies, or, is the difference of $635 million due to chance?
- If there is a difference, what is it?

To answer these questions, perform a two-sample t-test. A t-test lets you use data from a sample to make inferences about the larger population.

To perform the t-test, click on the red triangle next to **Oneway Analysis** and select **t Test**. Figure 4.24 shows the results that have been added to the report.

Figure 4.24 t Test Results



Since the p-value of 0.0004 is less than the significance level of 0.05, you can conclude that the difference you are seeing in average profits for the sample data is not due to chance alone. From this you conclude that the average profits for pharmaceutical companies are higher than the average profits for computer companies in the larger population.

Use the confidence interval limits to determine how much difference exists in the profits of both types of companies. Look at the **Upper CL Dif** and **Lower CL Dif** values shown in Figure 4.24. From these

values, you can conclude that the average for pharmaceutical companies is between $324 million and $944 million higher than the average for computer companies.

Comparing Proportions

If you have categorical *X* and *Y* variables, you can compare the proportions of the *Y* variable levels across the *X* variable levels.

Scenario

To illustrate comparing proportions, look at the financial data for 32 companies from the pharmaceutical and computer industries. Suppose you want to find out whether the populations of computer and pharmaceutical companies consist of the same proportions of small, medium, and big companies.

Assess the Relationship

To assess the relationship between company size and type, create a mosaic plot:

1 Open the Companies.jmp sample data table.

Note: If you still have the Companies.jmp data file open from the previous example, you might have rows that are excluded or hidden. To return the rows to the default state (all rows included and none hidden), select **Rows** > **Clear Row States**.

---

2 Select **Analyze** > **Fit Y by X**.

3 Assign Size Co to the **Y, Response** role.

4 Assign Type to the **X, Factor** role. Figure 4.25 shows the completed window.

5 Click **OK**. Figure 4.26 shows the initial results.

Figure 4.25 Completed Fit Y by X Window

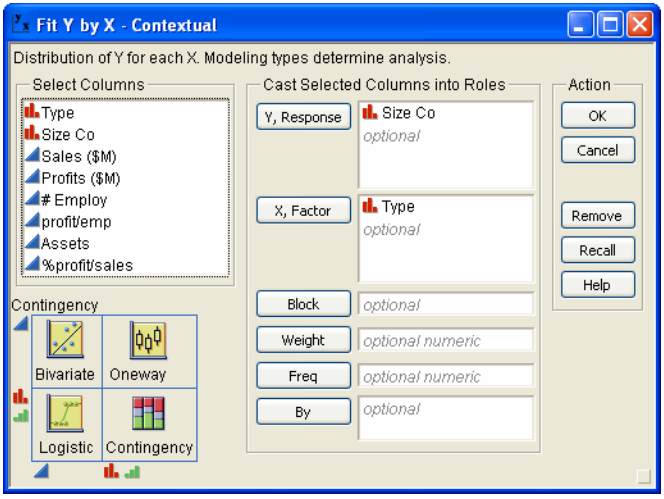


Figure 4.26 Company Size by Company Type

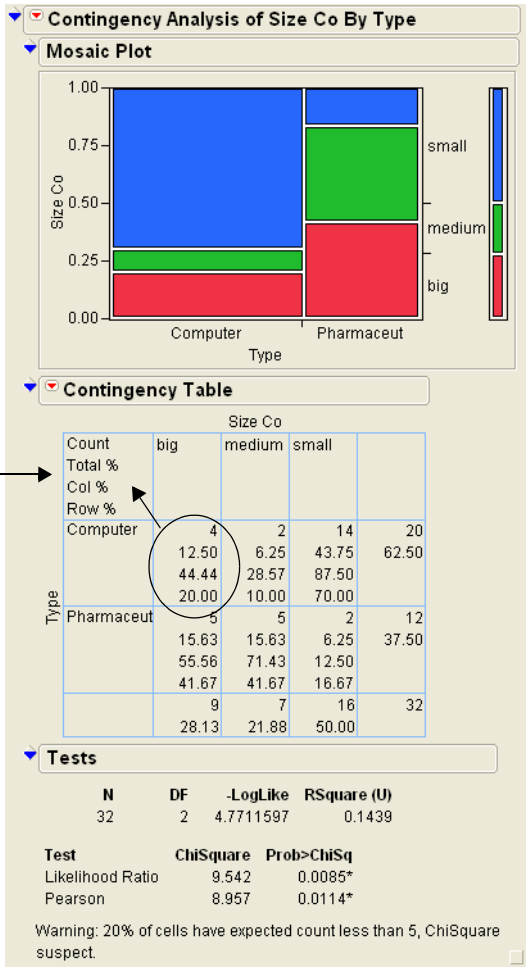
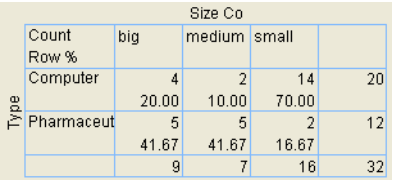


Table legend

The Contingency Table contains information that is not applicable for this example. To remove content, click the red triangle next to **Contingency Table**, and deselect **Total %** and **Col %**. Figure 4.27 shows the updated table.

Figure 4.27 Updated Contingency Table



The statistics in the Contingency Table are graphically represented in the Mosaic Plot. For example, the Mosaic Plot shows that the computer industry has a higher percentage of small companies compared to the pharmaceutical industry. The Contingency Table shows the exact statistics: 70% of computer

companies are small, and about 17% of pharmaceutical companies are small. You can use the Mosaic Plot and the Contingency Table to compare the percentages of medium and big companies between the two industries.

Interpreting the Test

Because you looked at only a sample of companies, you need to ask this question: Do the percentages differ in the broader populations of all computer and pharmaceutical companies, or, are the observed differences due to chance?

To answer this question, you can use the p-value from the Pearson test in the **Tests** outline. Since the p-value of 0.011 is less than the significance level of 0.05, you can conclude that the differences you are seeing in the sample data are not due to chance alone, and that the percentages differ in the broader population.

Using Regression with Multiple Predictors

The section "Using Regression with One Predictor," p. 92 showed you how to build simple regression models consisting of one predictor variable and one response variable. You can use *multiple regression* to predict the average response variable using two or more predictor variables. You can build and interpret multiple regression models as follows:

Scenario

To build a multiple regression model, look at sample data containing nutrition information for candy bars. Suppose you want to build a model to predict calories using these variables:

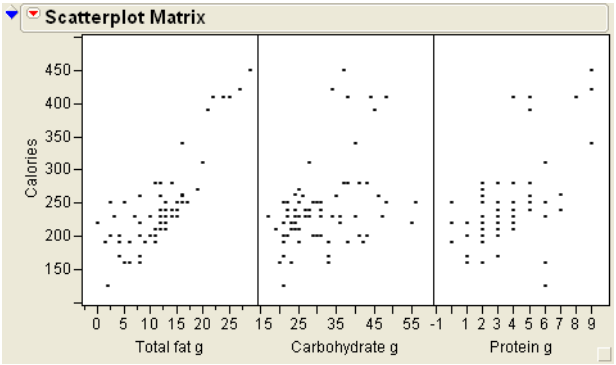
- Total fat
- Carbohydrates
- Protein

Assess the Relationship

To assess the relationship between calories and total fat, carbohydrates, and protein, create a scatterplot matrix:

1 Open the Candy Bars.jmp sample data table.
2 Select **Graph** > **Scatterplot Matrix**.
3 Assign Calories to the **Y, Columns** role.
4 Assign Total fat g, Carbohydrate g, and Protein g to the **X** role.
5 Click **OK**. Figure 4.28 shows the results.

Figure 4.28 Scatterplot Matrix Results



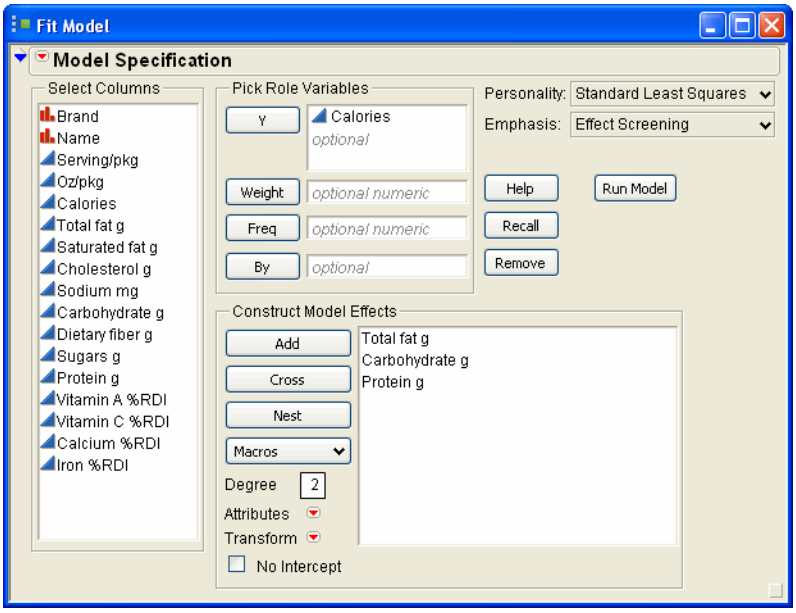
The scatterplot matrix shows that there is a positive correlation between calories and all three variables. The correlation between calories and total fat is the strongest. Now that you know there is a relationship, you can build a multiple regression model to predict average calories.

Build the Multiple Regression Model

To build the model, follow these steps:

1 Open the Candy Bars.jmp sample data table.
2 Select **Analyze** > **Fit Model**.
3 Assign Calories to the **Y** role.
4 Select Total Fat g and click **Add**. This assigns Total Fat g as a model effect.
5 Select Carbohydrate g and click **Add**.
6 Select Protein g and click **Add**.
7 Next to Emphasis, select **Effect Screening**. Figure 4.29 shows the completed Fit Model window.
8 Click the **Run Model** button.

Figure 4.29 Fit Model Window



The report window shows the model results. To interpret the model results, focus on these areas:

- "Actual by Predicted Plot," p. 104
- "Parameter Estimates," p. 105
- "Prediction Profiler," p. 105

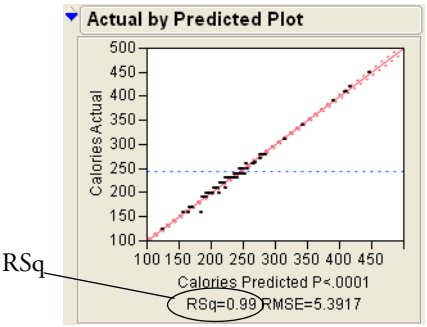
Note: For complete details about all of the model results, see *JMP Statistics and Graphics Guide*.

Actual by Predicted Plot

The Actual by Predicted Plot shows the actual calories versus the predicted calories. As the predicted values become closer to the actual values, the points on the scatterplot fall closer around the red line. See Figure 4.30. You can see that the model is predicting well.

Another measure of model accuracy is the RSq value (shown below the plot in Figure 4.30). The RSq value measures the percentage of variability in calories, as explained by the model. A value closer to 1 means a model is predicting well. In this example, the RSq value is 0.99.

Figure 4.30 Actual by Predicted Plot

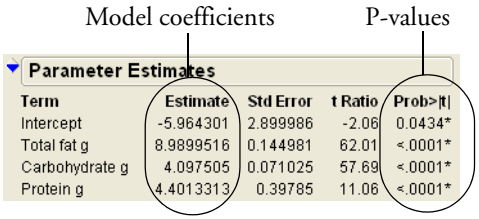


Parameter Estimates

The Parameter Estimates outline shows the following information:

- The model coefficients
- P-values for each parameter

Figure 4.31 Parameter Estimates Outline



In this example, the p-values are all very small (<.0001). This indicates that all three effects (fat, carbohydrate, and protein) contribute significantly when predicting calories.

You can use the model coefficients to predict the value of calories for particular values of fat, carbohydrate, and protein. Suppose you want to predict the average calories for any candy bar that has these characteristics:

- Fat = 11 g
- Carbohydrate = 43 g
- Protein = 2 g

Using these values, you can calculate the predicted average calories as follows:

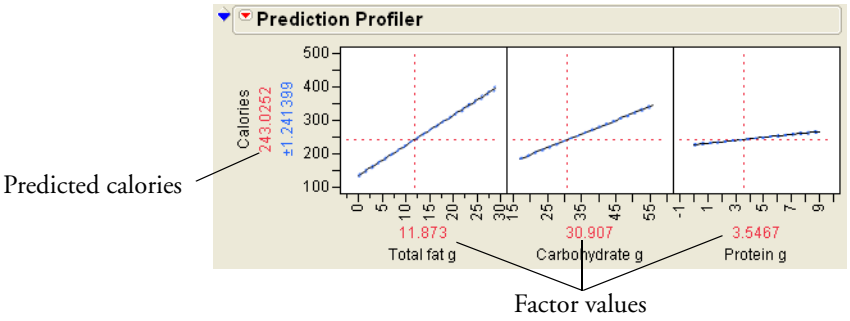
277.92 = -5.9643 + 8.99*11 + 4.0975*43 + 4.4013*2

The characteristics in this example are the same as the Milky Way candy bar (on row 59 of the data table). The actual calories for the Milky Way are 280, showing that the model predicts well.

Prediction Profiler

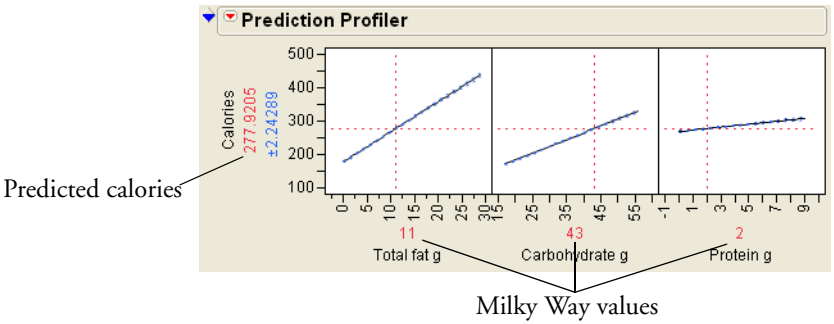
You can use the Prediction Profiler to assess how changes in the factors impact the predicted values. The profile lines show the magnitude of change in calories as the factor changes. The line for Total fat g is the steepest, meaning that changes in total fat have the largest impact on calories. See Figure 4.32.

Figure 4.32 Prediction Profiler



Click and drag the vertical line for each factor to see how the predicted value changes. Additionally, you can click the current factor values and change them. For example, click on the factor values and type the values for the Milky Way candy bar. See Figure 4.33.

Figure 4.33 Factor Values for the Milky Way



Note: For complete details about the Prediction Profiler, see the *JMP Statistics and Graphics Guide*.

Comparing Averages for Multiple Variables

The section "Comparing Averages for One Variable," p. 97, compared averages across the levels of a categorical variable. To compare averages across levels of two or more variables at once, use the *Analysis of Variance* technique (or ANOVA).

Scenario

To compare data across multiple variables, look at sample financial data within 32 pharmaceutical and computer companies. Suppose you want to compare company profits by these two variables:

- Type (pharmaceutical or computer)
- Size (small, medium, big)

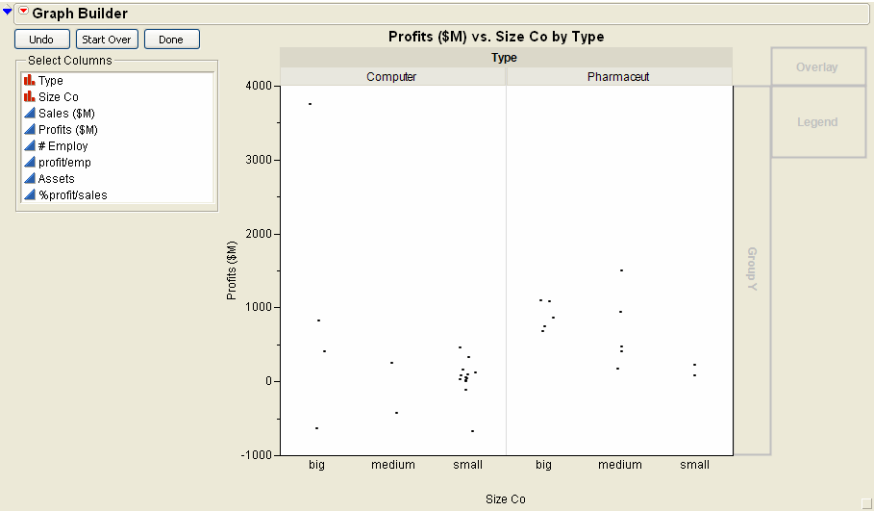
Assess the Relationship

To visualize the differences in profit for all combinations of type and size, create a graph:

1 Open the Companies.jmp sample data table.

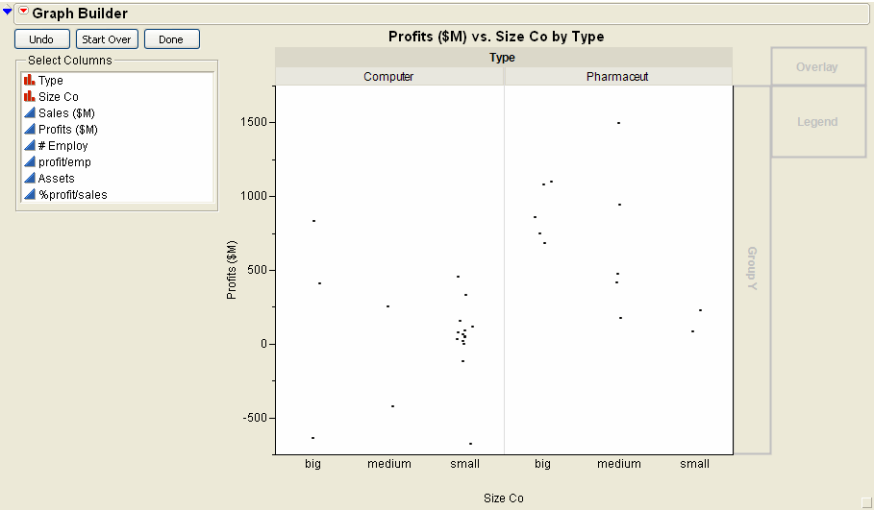
2 Select **Graph** > **Graph Builder**. The Graph Builder window appears.
3 Assign Profits ($M) to the **Y** zone.
4 Assign Size Co to the **X** zone.
5 Assign Type to the **Group X** zone. Figure 4.34 shows the completed graph.

Figure 4.34 Graph of Company Profits



You can see that one big computer company has very large profits. That outlier is stretching the scale of the graph, making it difficult to compare the other data points. Click on the outlier to select it, and then select **Rows** > **Exclude/Unexclude**. The point is removed, and the scale of the graph automatically updates. Figure 4.35 shows the updated graph.

Figure 4.35 Graph with Outlier Removed

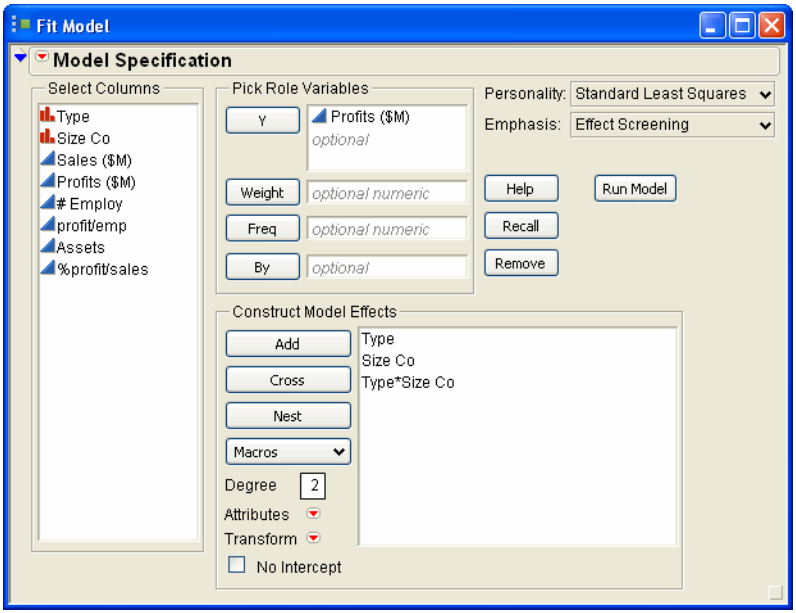


You can see that pharmaceutical companies have higher average profits. You can also see that for the pharmaceutical companies only, there appears to be a difference in the profits between company sizes. When the effect of one variable (company size) changes for different levels of another variable (company type), this is called an *interaction*.

Since this data is only a sample, you need to find out if the differences are limited to this sample (due to chance), or if the same patterns exist in the larger population of companies. To help you make this determination, perform the statistical tests using the **Fit Model** platform:

1 Select **Analyze** > **Fit Model**.
2 Assign Profits ($M) to the **Y** role.
3 Select both Type and Size Co.
4 Click the **Macros** arrow and select **Full Factorial**.
5 Next to Emphasis, select **Effect Screening**. Figure 4.29 shows the completed Fit Model window.
6 Click **Run Model**. The report window shows the model results.

Figure 4.36 Completed Fit Model Window



To decide whether the differences are real, or due to chance, examine the **Effect Tests** outline.

Note: For complete details about all of the **Fit Model** results, see the *JMP Statistics and Graphics Guide*.

Effect Tests

The Effect Tests outline (see Figure 4.37) shows the results of the statistical tests. There is a test for each of the effects included in the model on the Fit Model window (Type, Size Co, and Type*Size Co).

Figure 4.37 Effect Tests Outline

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Type	1	1	1401847.4	10.1368	0.0039*
Size Co	2	2	724616.2	2.6198	0.0927
Type*Size Co	2	2	448061.5	1.6200	0.2180

First look at the test for the interaction, the Type*Size Co effect. The p-value of 0.218 is large (greater than the significance level of 0.05). You previously observed that the pharmaceutical companies appeared to have different profits between company sizes. However, using this data, the effect test indicates that there is no interaction between type and size as it relates to profit. You can conclude that the differences do not apply to the broader population. Therefore, remove that effect from the model, and re-run the model, as follows:

1 Return to the Fit Model window.
2 In the Construct Model Effects box, select the **Type*Size Co** effect and click **Remove**.
3 Click **Run Model**. Figure 4.38 shows the new Effect Tests outline.

Figure 4.38 Effect Tests Outline

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Type	1	1	1356297.9	9.3768	0.0049*
Size Co	2	2	434161.3	1.5008	0.2410

The p-value for the Size Co effect is large, indicating that there are no differences between different sized companies in the broader population of all companies. The p-value for the Type effect is small, indicating that the differences you saw in the data between computer and pharmaceutical companies is not due to chance. You can conclude that there is a real difference between computer and pharmaceutical companies in the broader population.

About Advanced Modeling and Analysis

You can perform advanced statistical modeling and analysis using JMP platforms in these categories:

- "Modeling," p. 109
- "Multivariate Methods," p. 110
- "Reliability and Survival," p. 111

Note: You can access these platforms from the **Analyze** menu.

Modeling

Platforms that provide advanced modeling techniques include the following:

Screening Use the **Screening** platform to analyze screening designs when you want to sift through a large number of factors.

Nonlinear Use the **Nonlinear** platform to fit models that are nonlinear in the parameters.

Neural Net Use the **Neural Net** platform to fit a simple, one-layer neural net. A neural net model is efficient and flexible in modeling a wide variety of response surfaces.

Gaussian Process Use the **Gaussian Process** platform if you want a flexible response surface that responds to local features. You can also fit no-error models, which are common in areas like computer simulations, where a given input always results in the same output.

Partition Use the **Partition** platform to fit classification and regression trees.

Time Series Use the **Time Series** platform to analyze time-correlated data and to fit ARIMA and smoothing models.

Categorical Use the **Categorical** platform to tabulate categorical response data and to calculate test statistics. You can use this platform to analyze surveys and other categorical response data.

Choice Use the **Choice** platform to model complex choices, such as those used in consumer research experiments. You can use this platform for market research and product design.

Multivariate Methods

Platforms that explore relationships among multiple variables include the following:

Multivariate Use the **Multivariate** platform to investigate relationships among correlated response variables using these methods (to name a few):

- Scatterplot matrices
- Principal components
- Multivariate outlier plots

Cluster Use the **Cluster** platform to cluster observations using these methods:

- Hierarchical
- K-means
- EM (expectation maximization)

Principal Components Use the **Principal Components** platform to derive linear combinations of variables that explain the variability in the data.

Discriminant Use the **Discriminant** platform to derive linear combinations that you can use to predict group membership.

PLS Use the **PLS** platform to fit models using partial least squares. Use **PLS** when there are more x-variables than observations.

Item Analysis Use the **Item Analysis** platform to fit response curves from Item Response Theory, allowing you to analyze surveys or test data.

Reliability and Survival

Platforms that analyze and fit models to survival, reliability, and other time-to-event data include the following:

Life Distribution Use the **Life Distribution** platform to analyze and fit distributions to time-to-event data.

Fit Life by X Use the **Fit Life by X** platform to analyze data from accelerated life testing.

Recurrence Use the **Recurrence** platform to analyze time-to-event data where an event can occur more than once per unit. You can use this platform when a unit breaks down, is repaired, and is put back into service.

Survival Use the **Survival** platform to analyze survival data using product-limit (Kaplan Meier) computations.

Fit Parametric Survival Use the **Fit Parametric Survival** platform to fit parametric censored data. This platform is a personality of the **Fit Model** platform.

Fit Proportional Hazards Use the **Fit Proportional Hazards** platform to fit proportional hazard regression models using a Cox model. This platform is a personality of the **Fit Model** platform.

Chapter 5

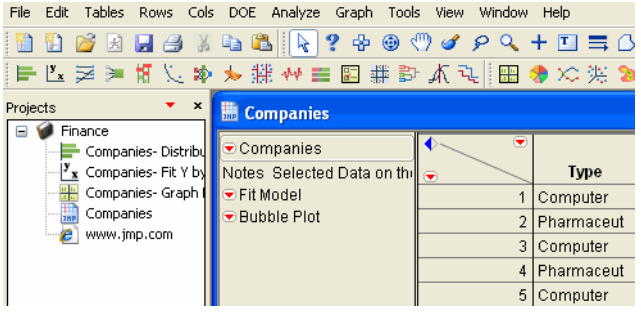
Customizing JMP

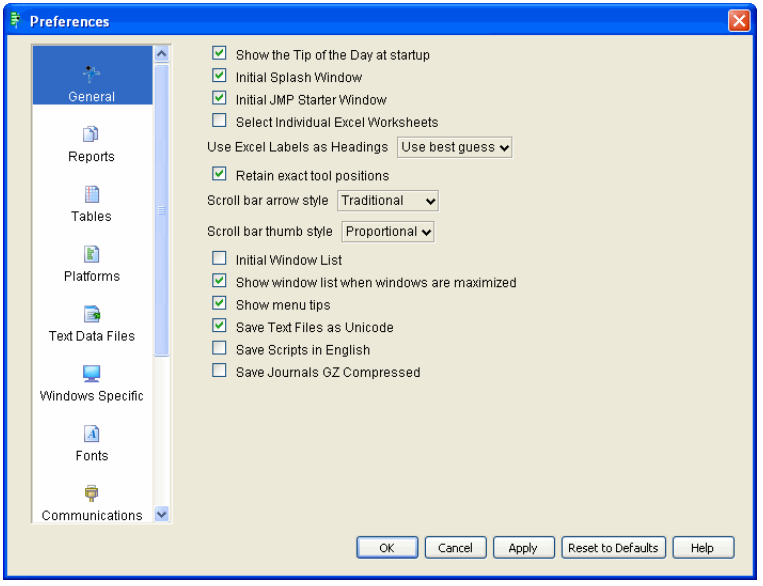
Saving Results, Table Variables, Scripts, and Preferences

Some of the ways you can customize JMP include the following actions:

- Saving platform results as journals or projects
- Creating table variables to use in formulas or as notes
- Creating scripts to reproduce analyses or graphs
- Changing preferences

Figure 5.1 Customization Examples





Contents

Saving Platform Results

You can save platform results using journals or projects. You can share these results and recreate them at any time. For full details about journals and projects, see the *JMP User Guide*.

Journals

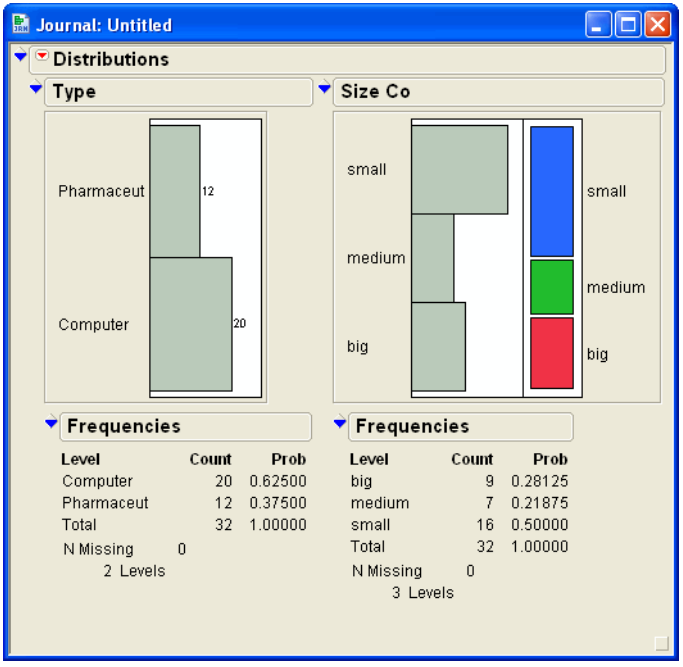
You can save platform reports for future viewing by creating a journal of the report window. The journal is a copy of the report window. You can edit or append other reports to a journal. The journal is not connected to the data table.

Example: Creating a Journal

As an example of creating a journal, perform the following steps:

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Distribution**.
3 Assign both Type and Size Co to the **Y, Columns** role.
4 Click **OK**.
5 On the **Type** outline, click on the red triangle menu and select **Histogram Options** > **Show Counts**.
6 On the **Size Co** outline, click on the red triangle menu and select **Mosaic Plot**.
7 To journal these results, select **Edit** > **Journal**. The results are duplicated in a journal window.

Figure 5.2 Journal of Distribution Results



The results in the journal are not connected to the data table. For example, if you click the Computer bar on the Type bar chart, no rows are selected in the data table.

Adding Additional Analyses

If you perform a different analysis, you can add the results to an existing journal. For example, perform these steps:

1 With a journal open, select **Analyze** > **Distribution**.
2 Assign profit/emp to the **Y, Columns** role.
3 Click **OK**.
4 Select **Edit** > **Journal**. The results are appended to the bottom of the journal.

Projects

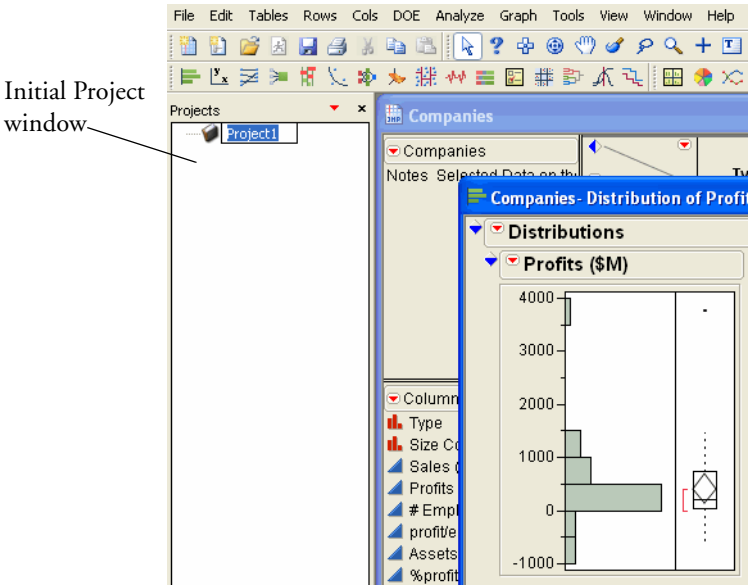
You can save multiple JMP file types (such as data tables, reports, journals, and scripts) into a single file by creating a project. The project file contains all the information needed to re-open all of the included files.

Example: Creating a Project

To create a project, perform these steps:

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Distribution**.
3 Assign Profits ($M) and profit/emp to the Y, Columns role.
4 Click **OK**.
5 To start a new project, select **File** > **New** > **Project**. The Project window appears on the left side of the JMP window. See Figure 5.3.

Figure 5.3 Initial Project Window

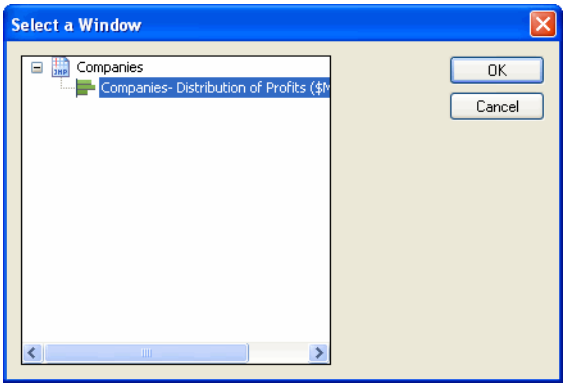


6 Type a name for the project.

7 To add the Distribution results to the project, right-click on the project name and select **Add Window**.

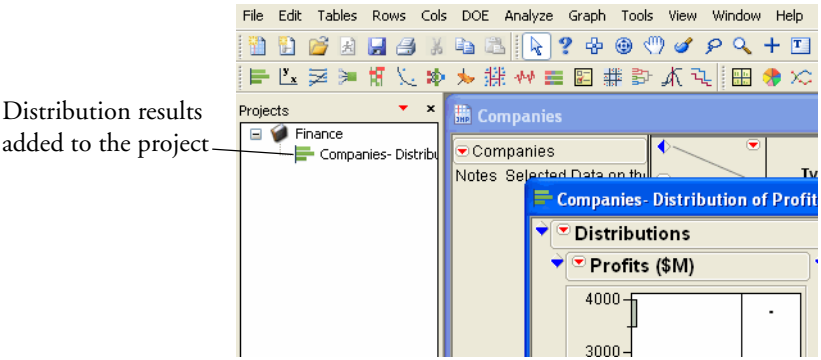
8 On the Select a Window window, select the Distribution results. See Figure 5.4.

Figure 5.4 Select the Distribution Results



9 Click **OK**. The Distribution results are added to the project. See Figure 5.5.

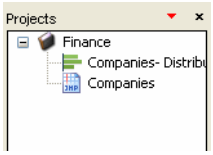
Figure 5.5 Distribution Results Added to the Project



10 To add the Companies.jmp data table to the project, repeat step 7 and select the Companies data table from the window.

11 Click OK. The data table is added to the project. See Figure 5.6.

Figure 5.6 Companies Data Table Added to the Project



You can double-click on the links in the project to open the data table and recreate the Distribution results.

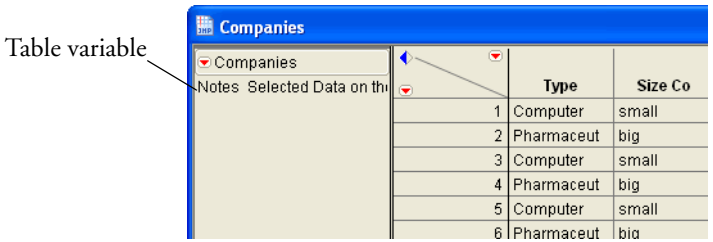
Working with Table Variables

You can use table variables as text to attach notes to a data table. You can also use table variables as numbers in column formulas.

Example: Viewing and Editing Table Variables

To see an example of a table variable, open the Companies.jmp sample data table. A table variable called **Notes** appears in the table panel. See Figure 5.7.

Figure 5.7 Table Variable



To view or edit the contents of the table variable, double-click on the contents portion. See Figure 5.8. If you double-click on the table variable name (in this example, Notes), you can edit the name or contents of the table variable. See Figure 5.9.

Figure 5.8 View the Contents of a Table Variable

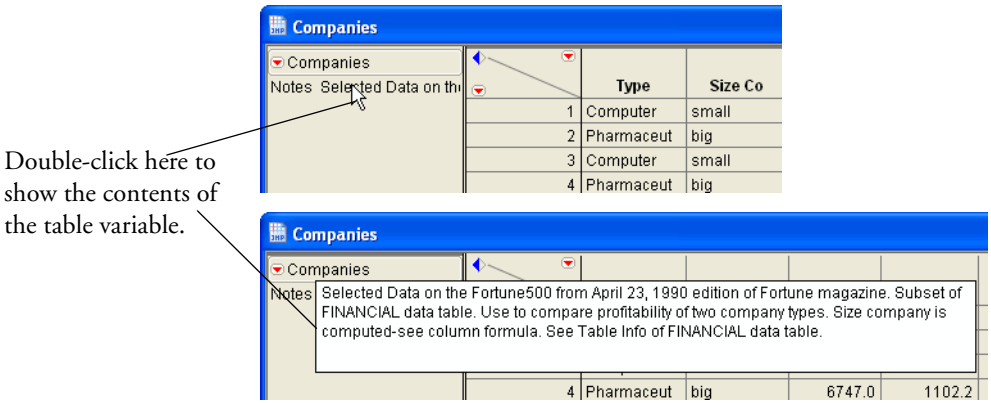
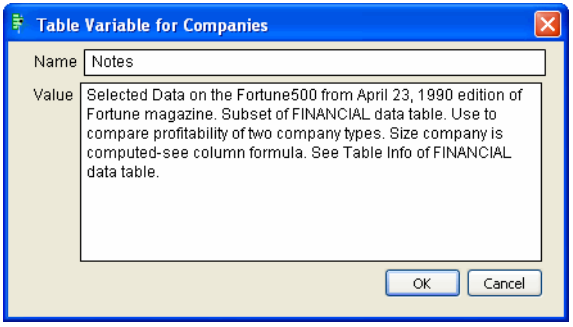


Figure 5.9 Table Variable Window

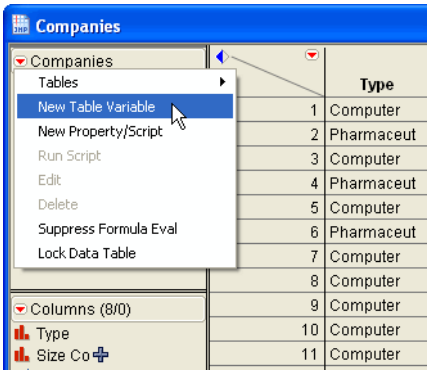


Creating Table Variables

To create a table variable, perform these steps:

1 Open the Companies.jmp sample data table.
2 Click on the red triangle menu next to Companies and select **New Table Variable**. See Figure 5.10.

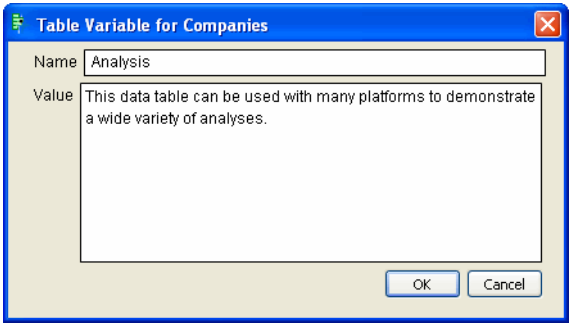
Figure 5.10 Create a New Table Variable



3 Type a name for the table variable in the **Name** box.

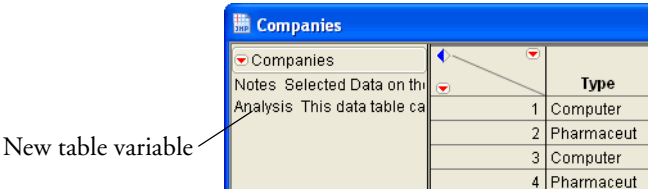
4 Type the contents of the variable in the **Value** box. See Figure 5.11.

Figure 5.11 Table Variable Window



5 Click **OK**. The table variable is added to the table panel. See Figure 5.12.

Figure 5.12 Table Variable is Added



Working with Scripts

While a platform report window is open, you can create scripts to reproduce your analyses or graphs. The scripts are saved to the data table, and you can run them at any time.

Example: Creating and Running a Script

To create and run a script, perform these steps:

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Distribution**.
3 Assign Type and profits/emp to the **Y, Columns** role.
4 Click **OK**.
5 Click on the red triangle menu next to Type and select the following:
   – **Histogram Options** > **Show Counts**
   – **Confidence Interval** > **0.95**
6 Click on the red triangle menu next to profit/emp and select the following:
   – **Outlier Box Plot**, to remove the outlier box plot
   – **CDF Plot**
7 Click on the red triangle menu next to Distribution and select **Stack**.
8 To save this analysis as a script, click on the red triangle menu next to Distributions and select **Script** > **Save Script to Data Table**. See Figure 5.13. The script appears in the table panel. See Figure 5.14.

Figure 5.13 Save Script to Data Table

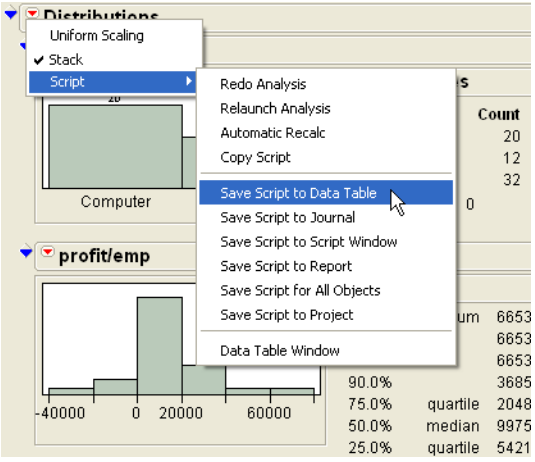
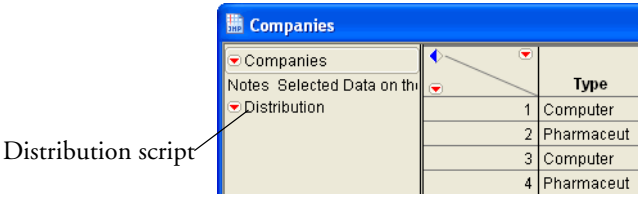


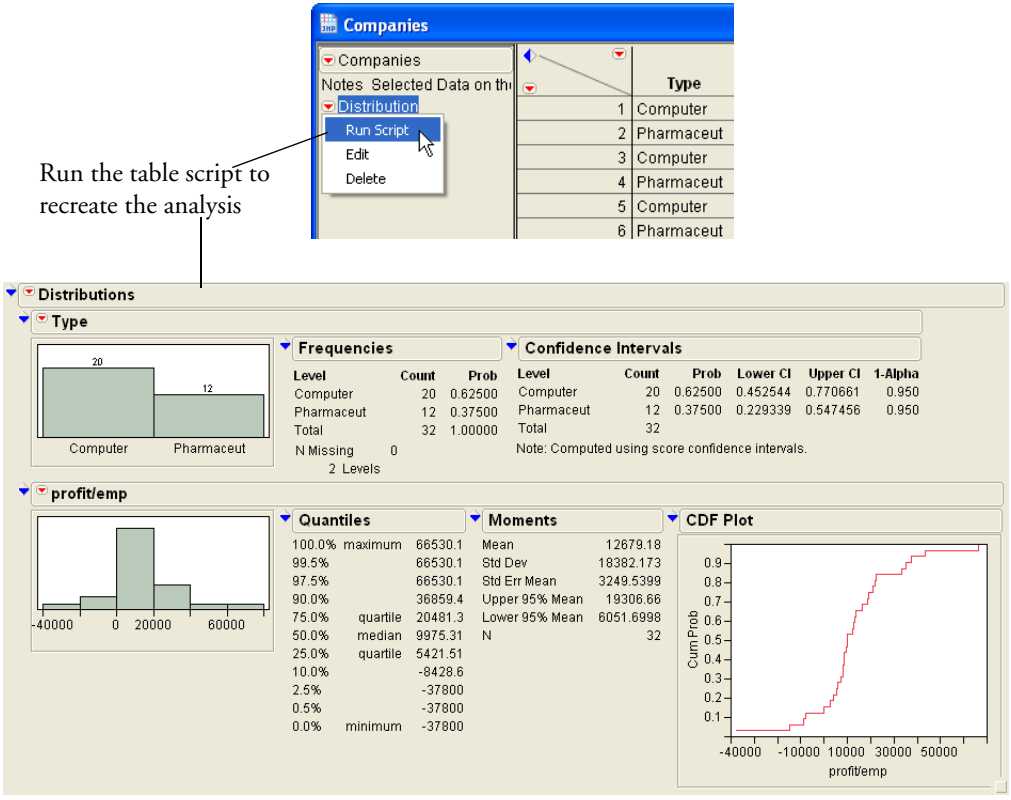
Figure 5.14 Distribution Script



9 Close the Distribution report window.

10 To recreate the analysis, click on the red triangle menu next to the script (Distribution) and select **Run Script**. See Figure 5.15.

Figure 5.15 Running the Distribution Script



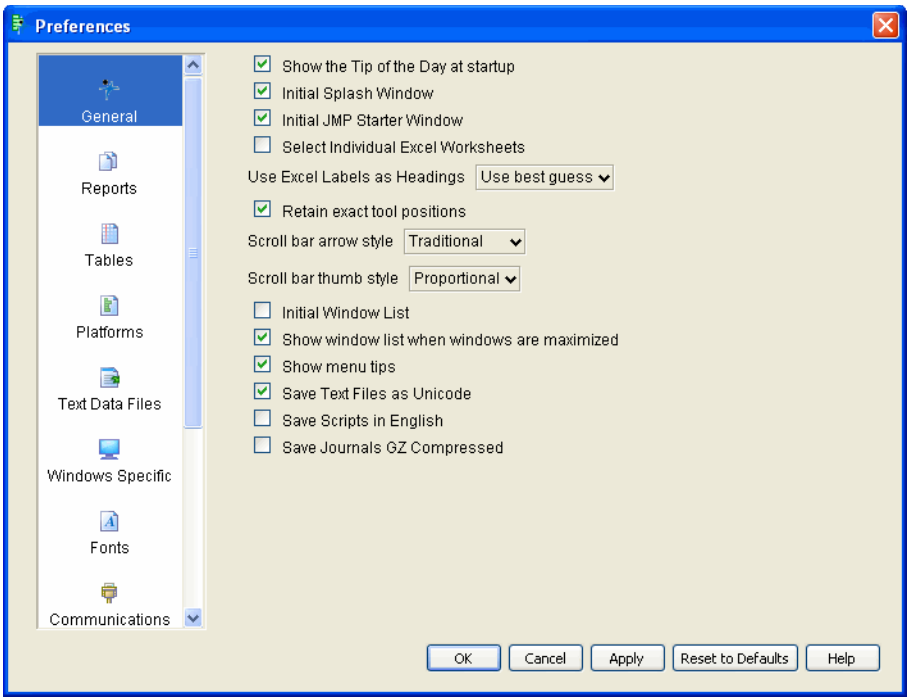
Scripts and JSL

The scripts you created in this section contain JMP Scripting Language (JSL) commands. You can develop JSL scripts independently of platforms and data tables. For complete details about JSL, see the *JMP Scripting Guide*.

Changing Preferences

You can change preferences in JMP using the Preferences window. To open the Preferences window (see Figure 5.16), select **File** > **Preferences**.

Figure 5.16 Preferences Window



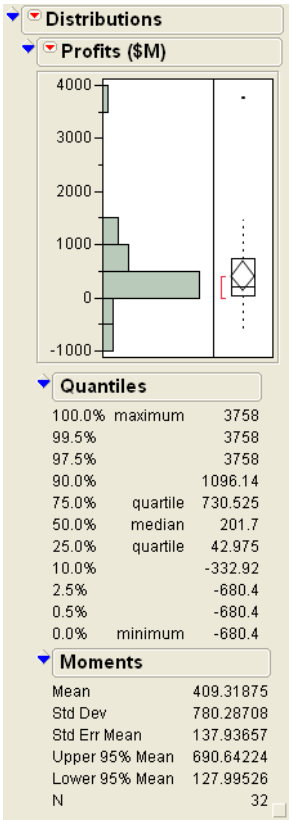
On the left side of the Preferences window is a list of categories. On the right side of the window are all of the preferences that you can change for the selected category.

Example: Changing Preferences

Every platform report window has options that you can turn on or off. However, your changes to these options are not remembered the next time you use the platform. If you want JMP to remember your changes every time you use the platform, change those options in the preferences window. For example, perform these steps:

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Distribution**.
3 Assign Profits ($M) to the **Y, Columns** role.
4 Click **OK**. Figure 5.17 shows the report window.

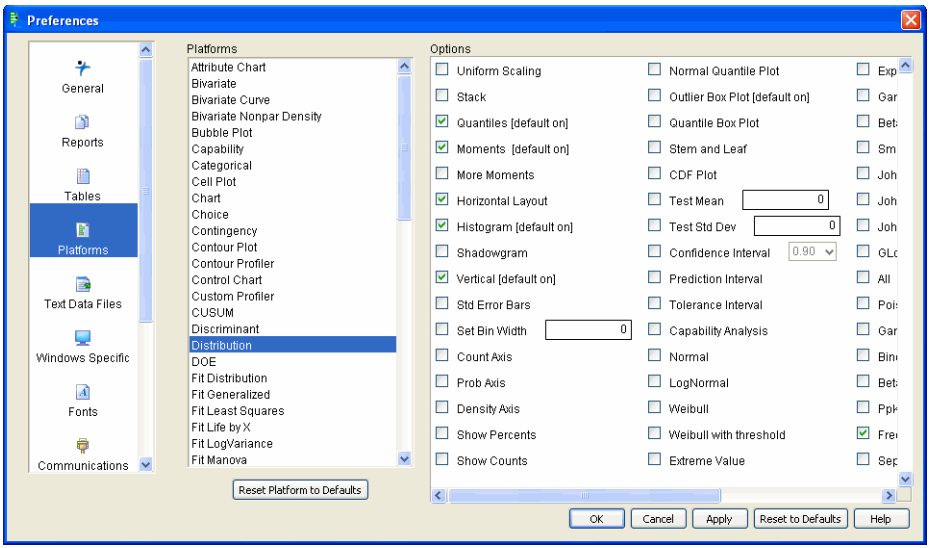
Figure 5.17 Distribution Report Window



Notice that the histogram is vertical, and the report window includes an outlier box plot. If you want to change the histogram to horizontal, and remove the outlier box, you can select the appropriate options on the Profits ($M) red triangle menu. However, if you want those preferences to always be in effect when you use the platform, then change them in the Preferences window:

1 Select **File** > **Preferences**.
2 Select the **Platforms** category in the left panel.
3 Select **Distribution** from the Platforms list.
4 Select the **Horizontal Layout** option.
5 Deselect the **Outlier Box Plot** option. Figure 5.18 shows the Preferences window.
6 Click **OK**.

Figure 5.18 Distribution Preferences



If you repeat the Distribution analysis, you can see that the histogram is now horizontal and the outlier box plot does not appear. These preferences remain the same until you change them.

For details about all of the preferences, see the *JMP User Guide*.

Chapter 6

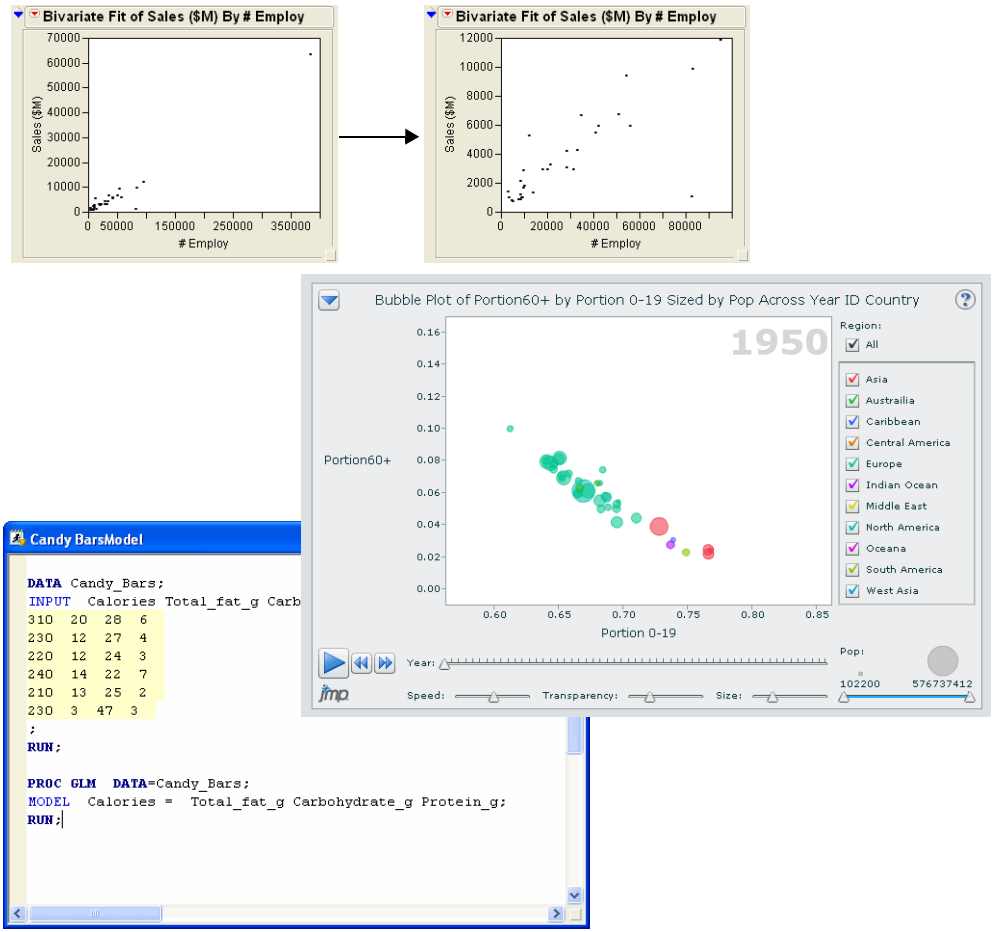
Special Features

Automatic Updating, Using Flash, and Integrating with SAS

Using some of the special features in JMP, you can do the following actions:

- Update analyses or graphs automatically
- Create files to show results outside of JMP using the Adobe Flash Player
- Integrate with SAS to access advanced analytical features

Figure 6.1 Examples of Special Features



Contents

Automatically Updating Analyses and Graphics

When you make a change to a data table, you can use the Automatic Recalc feature to automatically update analyses and graphs that are associated with the data table. For example, if you exclude, unexclude, or delete values in the data table, that change is instantly reflected in the associated analyses or graphs. Note the following information:

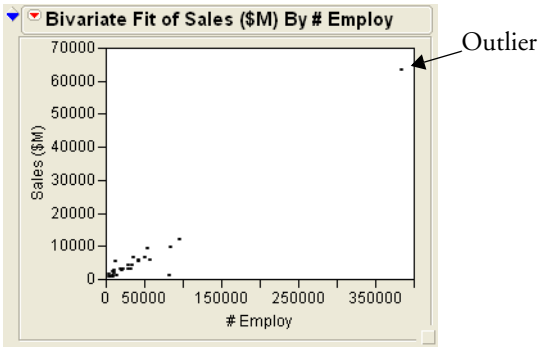
- Some platforms do not support Automatic Recalc. For more information, see the *JMP Statistics and Graphics Guide*.
- For the supported platforms in the **Analyze** menu, Automatic Recalc is turned off by default.
- For the supported platforms in the **Graph** menu, Automatic Recalc is turned on by default (except for Variability/Gauge Chart, Capability, and Control Chart).

Example: Using Automatic Recalc

To use Automatic Recalc, look at financial data for 32 companies from the pharmaceutical and computer industries.

1 Open the Companies.jmp sample data table.
2 Select **Analyze** > **Fit Y by X**.
3 Assign Sales ($M) to the **Y, Response** role.
4 Assign # Employ to the **X, Factor** role.
5 Click **OK**. See Figure 6.2.

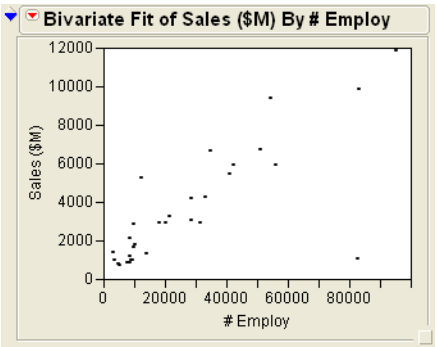
Figure 6.2 Initial Scatterplot



You can see from the initial scatterplot that one company has significantly more employees and sales than the other companies. You decide that this company is an outlier, and you want to exclude that point. Before you exclude the point, turn on Automatic Recalc so that your scatterplot is updated automatically when you make the change.

6 Turn on Automatic Recalc by clicking on the red triangle menu and selecting **Script** > **Automatic Recalc**.
7 Click on the outlier to select it.
8 Select **Rows** > **Exclude/Unexclude**. The point is excluded from the analysis and the scatterplot automatically updates. See Figure 6.3.

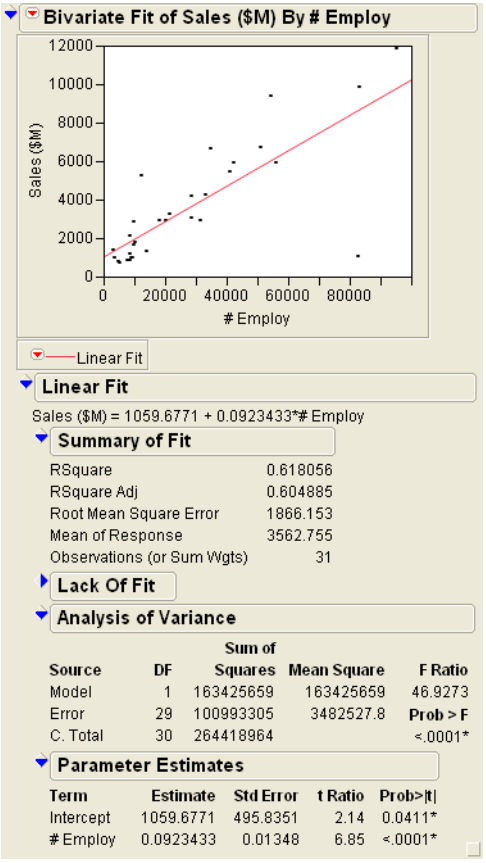
Figure 6.3 Updated Scatterplot



If you fit a regression line to the data, the point in the lower right corner has a significant influence on the slope of the line. If you then exclude that point with Automatic Recalc turned on, you see the slope of the line change.

9 Fit a regression line by clicking on the red triangle menu and selecting **Fit Line**. Figure 6.4 shows the regression line and analysis results added to the report window.

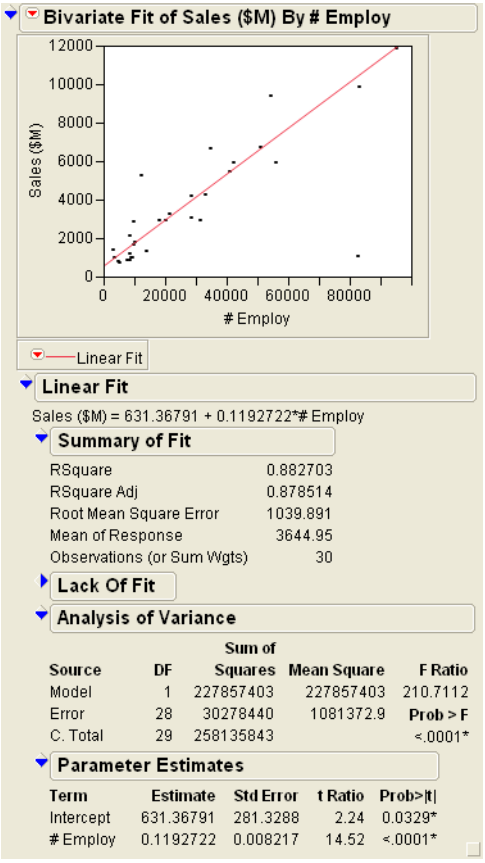
Figure 6.4 Regression Line and Analysis Results



10 Click on the point to select it.

11 Select **Rows** > **Exclude/Unexclude**. The regression line and analysis results are automatically updated, reflecting the exclusion of the point. See Figure 6.5.

Figure 6.5 Updated Regression Line and Analysis Results



When you exclude a point, it does not hide it in the scatterplot; it simply does not include it in the analysis. You can also hide the point in the scatterplot by clicking on the point and selecting **Rows** > **Hide/Unhide**.

Creating Flash Versions of the Profiler and Bubble Plot

If you want to show the interactive results of a profiler or bubble plot outside of JMP, you can export a .SWF file that can be viewed using the Adobe Flash Player. The .SWF file can be imported into presentations and web applications. You can also save the results as an HTML page with the .SWF output embedded.

To demonstrate creating a Flash version of the Bubble Plot, do the following steps:

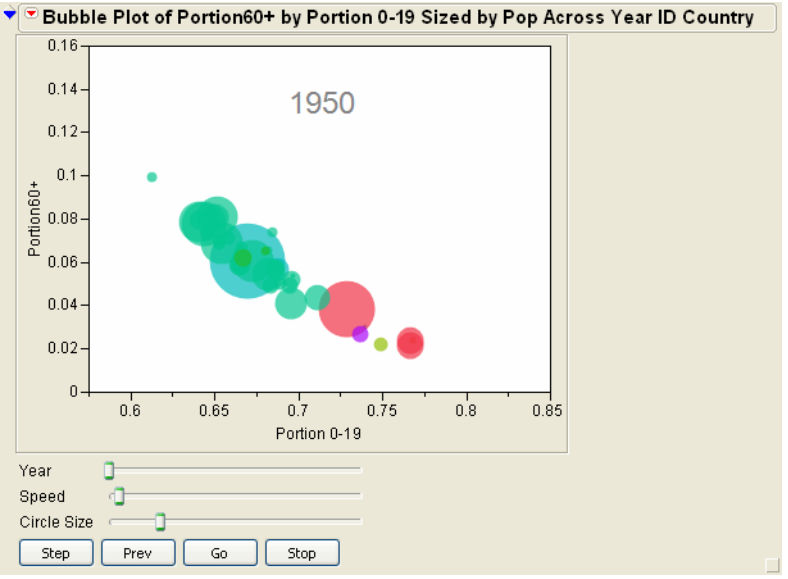
1 Open the PopAgeGroup.jmp sample data table.

2 Select **Graph** > **Bubble Plot**.

3 Assign Portion60+ to the **Y** role.

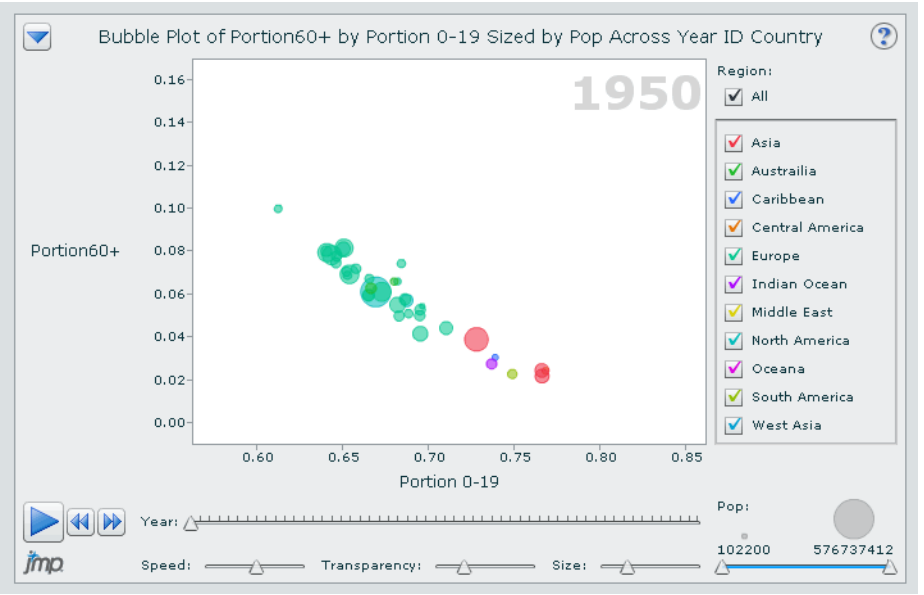
4 Assign Portion 0-19 to the **X** role.
5 Assign Country to the **ID** role.
6 Assign Year to the **Time** role.
7 Assign Pop to the **Sizes** role.
8 Assign Region to the **Coloring** role.
9 Click **OK**. See Figure 6.6.

Figure 6.6 Initial Bubble Plot



10 Click on the red triangle menu and select **Save As Flash (SWF)**.
11 In the Save As SWF window, select the location to which you want to save the file.
12 Click **Save**. The Flash version of the Bubble Plot is saved as HTML and appears in a web browser. See Figure 6.7.

Figure 6.7 Flash Version of Bubble Plot



For additional information about the Flash versions of the Profiler and Bubble Plot platforms, or for instructions on importing the Flash versions into Microsoft PowerPoint 2007, see the following link:

www.jmp.com/support/swfhelp/en/

Integrating JMP and SAS

Using JMP, you can interact with SAS as follows:

- You can write or create SAS code in JMP.
- You can submit SAS code and view the results in JMP.
- You can connect to a SAS Metadata Server or a SAS Server on a remote machine.
- You can connect to SAS on your local machine.
- You can open and browse SAS data sets.
- You can retrieve and view data sets generated by SAS.

Note: For complete details about integrating JMP and SAS, see the *JMP User Guide*.

As an example of integrating JMP and SAS, perform the steps in the following sections:

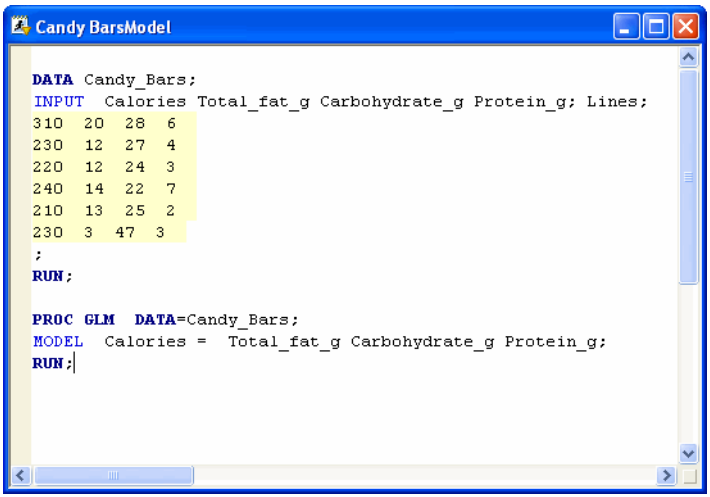
- "Create SAS Code," p. 135
- "Submit SAS Code," p. 135

Create SAS Code

To create SAS code in JMP, look at nutrition data for candy bars.

1 Open the Candy Bars.jmp sample data table.
2 Select **Analyze** > **Fit Model**.
3 Assign Calories to the **Y** role.
4 Assign Total fat g, Carbohydrates g, and Protein g as model effects.
5 To generate the SAS code that lets you run the same model in SAS, click on the red triangle menu and select **Create SAS Job**. Figure 6.8 shows the SAS code (not all of the data is shown here.)

Figure 6.8 SAS Code

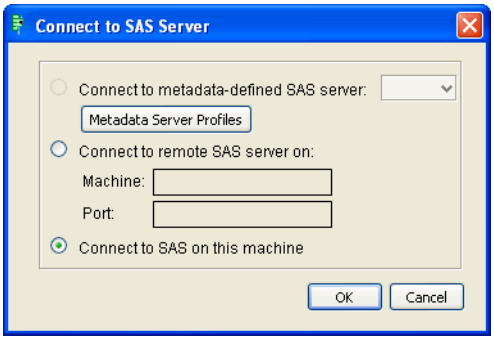


Submit SAS Code

To submit code (see Figure 6.8) to SAS from within JMP, take the following steps:

1 Perform steps 1 through 4 in the section "Create SAS Code," p. 135. Step 5 is optional.
2 In the Fit Model window, click on the red triangle menu and select **Submit to SAS**.
3 In the **Connect to SAS Server** window (see Figure 6.9), choose a method to connect to SAS (if you are not already connected). For this example, select **Connect to SAS on this machine**.

Figure 6.9 Connect to SAS Server



JMP connects to SAS. SAS runs the model and sends the results back to JMP. The results can appear as SAS output, HTML, RTF, PDF, or JMP report format (you can choose the format using JMP Preferences). Figure 6.10 shows the results formatted as a JMP report. See the *JMP User Guide* for details.

Figure 6.10 SAS Results Formatted as a JMP Report

Report: SAS Results

The SAS System
The GLM Procedure

Number of Observations

Number of Observations Read	75
Number of Observations Used	75

Dependent Variable: Calories

Overall ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	282358	94119.3	3237.58	7.7e-76
Error	71	2064.03	29.0709	.	.
Corrected Total	74	284422	.	.	.

Fit Statistics

R-Square	Coeff Var	Root MSE	Calories Mean
0.99274	2.21858	5.39174	243.027

Type I Model ANOVA

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Total_fat_g	1	185260	185260	6372.68	3e-71
Carbohydrate_g	1	93540.4	93540.4	3217.67	7e-61
Protein_g	1	3557.86	3557.86	122.386	4.2e-17

Type III Model ANOVA

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Total_fat_g	1	111777	111777	3844.97	1.4e-63
Carbohydrate_g	1	96756.1	96756.1	3328.28	2.2e-61
Protein_g	1	3557.86	3557.86	122.386	4.2e-17

Solution

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-5.9643	2.89999	-2.0567	0.04339
Total_fat_g	8.98995	0.14498	62.0078	1.4e-63
Carbohydrate_g	4.0975	0.07102	57.6913	2.2e-61
Protein_g	4.40133	0.39785	11.0628	4.2e-17

Appendix A

Navigating JMP

Menus and Shortcuts

This appendix describes the platforms on the **Analyze**, **Graph**, and **Tables** menus, and describes common keyboard shortcuts and mouse actions.

Note: The other menus are described in the *JMP Design of Experiments Guide* and the *JMP User Guide*.

Contents

The Analyze Menu

The following table describes all of the platforms in the **Analyze** menu.

Table A.1 Platform Descriptions in the Analyze Menu

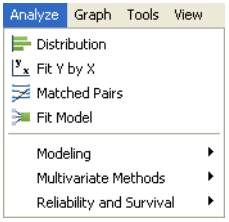
Platform	Description	Menu
Distribution	Provides a histogram for continuous data and a bar chart for nominal or ordinal data, along with relevant summary statistics. Presents options for many one-sample analyses, based on modeling type.	
Fit Y by X	Shows plots that describe the relationship between any two variables. Provides analyzes based on the modeling types of the two variables, such as bivariate, one-way, logistic, and contingency analysis.	
Matched Pairs	Analyzes two continuous variables that are measurements on the same experimental unit or subject.	
Fit Model	Fits models involving one or more Y variables and multiple X variables. Techniques include standard least squares, stepwise, MANOVA, loglinear variance, logistic, proportional hazards, parametric survival, and generalized linear models.	

Table A.1 Platform Descriptions in the Analyze Menu

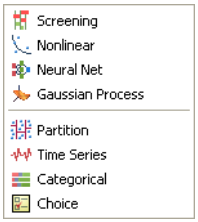
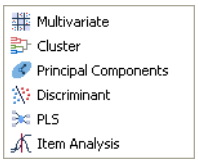
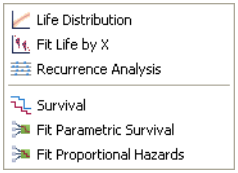
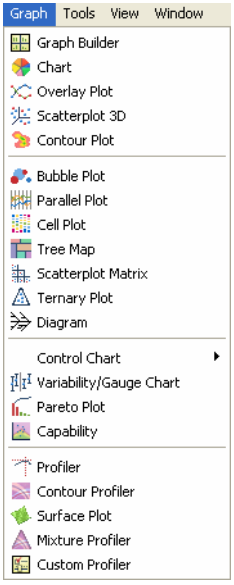
Platform	Description	Menu
Modeling	• Screening helps in the analysis of two-level designs by showing which effects are large. • Nonlinear fits models that are nonlinear in their parameters. • Neural Net fits a simple one-layer neural net model. • Gaussian Process fits no-error-term models that can perfectly interpolate the data. • Partition recursively partitions the data to fit a decision tree. • Time Series lets you explore and forecast univariate time-series data. • Categorical tabulates and summarizes categorical response data. • Choice lets you analyze and model complex choices, such as those used in consumer research experiments.	
Multivariate Methods	• Multivariate explores how multiple variables relate to each other. • Cluster lets you cluster (or group) the rows of a data table according to their similarities across multiple variables. • Principal Components derives linear combination of variables that capture as much of the variability in the original variables as possible. • Discriminant derives linear combinations that predict group membership. • PLS fits models using partial least squares. This is useful when there are more x-variables than observations. • Item Analysis fits response curves from Item Response Theory, enabling you to analyze surveys or test data.	

Table A.1 Platform Descriptions in the Analyze Menu

Platform	Description	Menu
Reliability and Survival	• Life Distribution analyzes and fits distributions to time-to-event data. • Fit Life by X analyzes data from accelerated life testing. • Recurrence Analysis analyzes time-to-event data where an event can occur more than once per unit. You can use this platform when a unit breaks down, is repaired, and is put back into service • Survival analyzes survival data using product-limit (Kaplan Meier) computations. • Fit Parametric Survival fits models to parametric censored data. • Fit Proportional Hazards fits proportional hazard regression models using a Cox model.	

The Graph Menu



The following table describes all of the platforms in the **Graph** menu.

Table A.2 Platform Descriptions in the Graph Menu

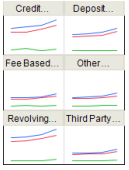
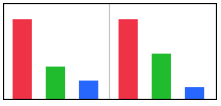
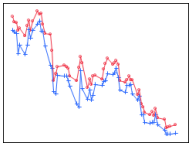
Platform	Description	Example
Graph Builder	Uses an interactive interface to create flexible graphs that are useful for visualizing multivariate data.	
Chart	Plots bar, line, pie, needle, and point charts of user-specified summary statistics for multiple Y variables across the values of up to two X variables.	
Overlay Plot	Plots one or more numeric Y variables across the values of a single X variable. Constructs several plots in one window using grouping variables.	

Table A.2 Platform Descriptions in the Graph Menu

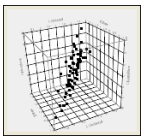
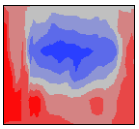
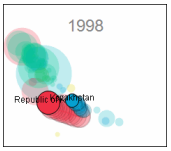
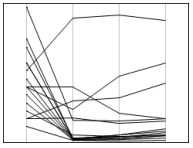
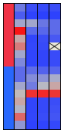
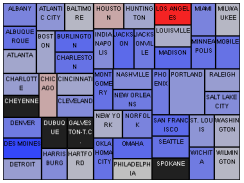
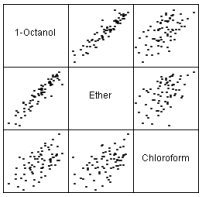
Platform	Description	Example
Scatterplot 3D	Produces three-dimensional rotatable plots for multiple Y variables. Also displays normal contour ellipsoids, nonparametric density contours, and biplot rays.	
Contour Plot	Constructs contour plots for one or more numeric response variables, Y, across a grid defined by two numeric X variables.	
Bubble Plot	Displays a scatterplot with points represented as bubbles. The bubbles can be grouped, sized, colored, and animated across time.	
Parallel Plot	Draws a coordinate plot that shows connected line segments representing each row of the data table.	
Cell Plot	Produces a rectangular array of cells drawn with a one-to-one correspondence to data table values, and colored by the values in the cells.	
Tree Map	Displays tree maps, which can be thought of as bar charts that have been folded over in two dimensions.	
Scatterplot Matrix	Displays scatterplots for all pairs of variables, or for a selection of Y variables plotted against one or more X variables.	

Table A.2 Platform Descriptions in the Graph Menu

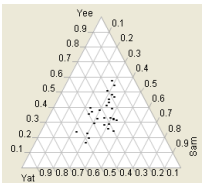
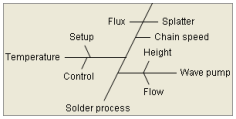
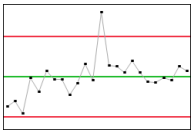
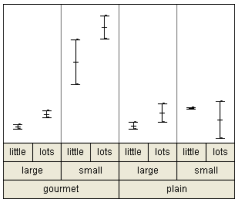
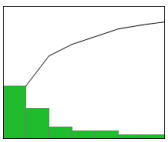
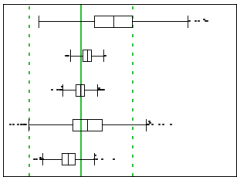
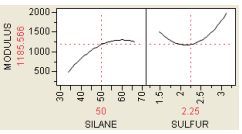
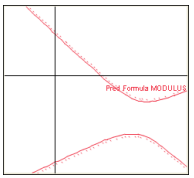
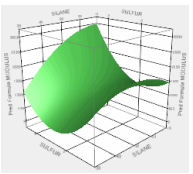
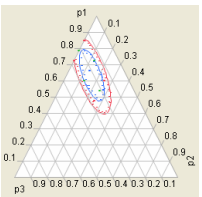
Platform	Description	Example
Ternary Plot	Displays the compositional makeup of mixture data on a two-dimensional plot.	
Diagram	Creates Fishbone (Ishikawa) diagrams	
Control Chart	Creates plots of time-ordered data, with control limits that bound the variation.	
Variability/Gauge Chart	Creates a graph of one Y variable versus multiple X variables. Analyzes Gauge R&R data and calculates variance components.	
Pareto Plot	Creates a bar chart that displays causes of defects in order of decreasing frequency.	
Capability	Conducts capability analysis on multiple responses.	
Profiler	Displays prediction traces for models. Allows optimization using desirability functions, and Monte-Carlo simulation.	

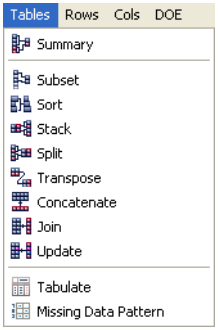
Table A.2 Platform Descriptions in the Graph Menu

Platform	Description	Example
Contour Profiler	Displays contours of a prediction formula.	
Surface Profiler	Creates a 3-dimensional, rotatable, surface plot of a model.	
Mixture Profiler	Displays contours for a prediction formula for models involving mixture factors.	
Custom Profiler	Provides an interface for simultaneous optimization of responses.	

The Tables Menu

The following table describes all of the platforms in the **Tables** menu.

Table A.3 Platform Descriptions in the Tables Menu

Platform	Description	Menu
Summary	Creates a table containing summary statistics for each level of grouping variables.	
Subset	Creates a subset table of the original table.	
Sort	Sorts a table by one or more columns.	
Stack	Stacks columns and creates a new table.	
Split	Splits one or more columns by the values in another column.	
Transpose	Creates a new table where the columns of the new table are the rows of the original table.	
Concatenate	Stacks multiple tables on top of each other.	
Join	Adds columns from a second table to the main table, by matching columns or row number.	
Update	Replaces data in a table with values from a second table.	
Tabulate	Constructs tables of summary statistics using an interactive interface.	
Missing Data Pattern	Describes the missing values in the table in terms of the columns where they are missing.	

Keyboard Shortcuts and Mouse Actions

This section describes frequently used keyboard shortcuts and mouse actions.

Files

The following table describes keyboard shortcuts for files.

Table A.4 Keyboard Shortcuts for Files

Command	Windows	Macintosh
New Data Table	Ctrl + N	Command + N
New Script	Ctrl + T	Shift + Command + N
New Journal	Ctrl + Shift + J	Option + Command + N
New Project	Ctrl + Shift + P	not applicable
Open File	Ctrl + O	Command + O
Save File	Ctrl + S	Command + S
Close File	Ctrl + W	Command + W
Exit JMP	Ctrl + Q	Command + Q
Print	Ctrl + P	Command + P
JMP Help	F1	Shift + Command + ?
JMP Preferences	Ctrl + K	Command + .
Run script	Ctrl + R	Command + R
Journal a report	Ctrl + J	Command + J

Data Tables

The following table describes general keyboard shortcuts for data.

Table A.5 General Keyboard Shortcuts for Data

Command	Windows	Macintosh
Cut	Ctrl + X	Command + X

Table A.5 General Keyboard Shortcuts for Data

Command	Windows	Macintosh
Paste	Ctrl + V	Command + V
Copy	Ctrl + C	Command + C
Copy as text	not applicable	Shift + Command + C
Undo	Ctrl + Z	Command + Z
Redo	Ctrl + Y	Shift + Command + Z
Select all	Ctrl + A	Command + A
Submit to SAS	F8	Command + Shift + R
Show previously selected row	F2	not applicable
Show next selected row	F3	not applicable
Scroll to a column	Double click the column name in the column panel	Double click the column name in the column panel
Increase font size	Ctrl + Shift + Plus Sign (+)	Command + Plus Sign (+)
Decrease font size	Ctrl + Shift + Minus Sign (-)	Command + Minus Sign (-)
Select where	Ctrl + Shift + W	not applicable
Exclude/Unexlcude	Ctrl + E	not applicable
Clear all selections from the data table	Esc	Esc

The following table describes keyboard shortcuts for searching, finding, and replacing data.

Table A.6 Keyboard Shortcuts for Searching, Finding, and Replacing Data

Command	Windows	Macintosh
Find	Ctrl + F	Command + F
Find Next	Ctrl + G	Command + G
Replace	Ctrl + H	Command + =
Replace and Find Next	Ctrl + I	Command + L

Windows

The following table describes keyboard shortcuts for working with windows.

Table A.7 Keyboard Shortcuts for Windows

Command	Windows	Macintosh
Close active window	Ctrl + W	Command + W
Show JMP Starter (also hide on Windows)	Ctrl + 0 (zero)	Command + 0 (zero)
Show Log window (also hide on Windows)	Ctrl + Shift + L	Command + 1
Move window to back	Ctrl + B	not applicable
Redraw window	Ctrl + D	Command + D

Mouse Actions

The following table describes frequently used mouse actions.

Table A.8 Frequently Used Mouse Actions

Action	Windows	Macintosh
Change the number of displayed decimal places in a report	Double-click the column in the report	Double-click the column in the report
Turn numeric report output into a data table	Right-click and select Make Into Data Table	Ctrl + Click and select Make Into Data Table
Display row legend in most graphics	Right-click in the graphic and select Row Legend	Ctrl + Click in the graphic and select Row Legend
Select points on a graph	Click and drag within a plot. Extend selection by holding down the Shift key	Click and drag within a plot. Extend selection by holding down the Shift key
In certain reports, apply command to all similar reports in that window	Ctrl + Click the red triangle then select the command	Command + Click the red triangle then select the command
Resize all similar graphs in a report	Ctrl + resize one graph	Command + resize one graph

Table A.8 Frequently Used Mouse Actions

Action	Windows	Macintosh
Create a subset of the data from a histogram	Select a bar or bars, and then double-click a bar; or, right-click and select Subset	Select a bar or bars, and then double-click a bar; or, right-click and select Subset
Select options (relating to journals)	Alt + right-click on blue triangle icon	not applicable
Show tree structure	Alt + right-click on blue triangle icon and select Show Tree Structure	Option + click on gray disclosure icon and select Edit > Show Tree Structure