



Version 9

# 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

**9.0.2**

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2010. *JMP*<sup>®</sup> 9 *Discovering JMP*. Cary, NC: SAS Institute Inc.

## **JMP<sup>®</sup> 9 Discovering JMP**

Copyright © 2010, SAS Institute Inc., Cary, NC, USA

ISBN 978-1-60764-600-6

All rights reserved. Produced in the United States of America.

**For a hard-copy book:** No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a Web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government Restricted Rights Notice:** Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

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

1st printing, September 2010

2nd printing, January 2011

3rd printing, June 2011

SAS<sup>®</sup> Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at [support.sas.com/publishing](http://support.sas.com/publishing) or call 1-800-727-3228.

JMP<sup>®</sup>, SAS<sup>®</sup> and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

## Get the Most from JMP<sup>®</sup>

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded Webcasts about how to get started with JMP
- video demos and Webcasts of new features and advanced techniques
- schedules for seminars being held in your area
- success stories showing how others use JMP
- a blog with tips, tricks, and stories from JMP staff
- a forum to discuss JMP with other users

**<http://www.jmp.com/getstarted/>**



# Contents

## Discovering JMP

---

<b>Gallery of JMP Graphs</b> .....	1
<b>About This Guide</b> .....	9
<b>1 Introducing JMP</b>	
<b>Basic Concepts</b> .....	11
Concepts You Should Know .....	13
How Do I Get Started? .....	13
Starting JMP .....	14
Using Sample Data .....	16
Understanding Data Tables .....	16
Understanding the JMP Workflow .....	18
Step 1: Launching a Platform and Viewing Results .....	19
Step 2: Removing the Box Plot .....	20
Step 3: Requesting Additional Output .....	21
Step 4: Interacting with Platform Results .....	21
How is JMP Different from Excel? .....	23
Formulas .....	23
Columns Names .....	23
Tables and Worksheets .....	23
The Data Grid .....	23
Analysis and Graph Reports .....	23
How Do I Find Help? .....	23
<b>2 Working with Your Data</b>	
<b>Preparing Your Data for Graphing and Analyzing</b> .....	25
Getting Your Data Into JMP .....	27
Copying and Pasting Data .....	27
Importing Data .....	27
Entering Data .....	29
Working with Data Tables .....	32
Editing Data .....	32

Selecting, Deselecting, and Finding Values .....	33
Viewing or Changing Column Information .....	37
Calculating Values With Formulas .....	39
Filtering Data .....	41
Managing Data .....	43
Requesting Summary Statistics .....	43
Creating Subsets .....	47
Joining Data Tables .....	49
Sorting Tables .....	51

### 3 Visualizing Your Data

<b>Graphing Data</b> .....	53
Looking at Single Variables .....	55
Histograms .....	55
Bar Charts .....	57
Comparing Multiple Variables .....	59
Scatterplots .....	60
Scatterplot Matrix .....	64
Side-by-Side Box Plots .....	67
Overlay Plots .....	69
Variability Chart .....	73
Graph Builder .....	75
Bubble Plots .....	80

### 4 Analyzing Your Data

<b>Distributions, Relationships, and Models</b> .....	85
About This Chapter .....	87
The Importance of Graphing Your Data .....	87
Understanding Modeling Types .....	89
Example: Modeling Type Results .....	90
Changing the Modeling Type .....	91
Analyzing Distributions .....	93
Distributions of Continuous Variables .....	93
Distributions of Categorical Variables .....	95
Analyzing Relationships .....	98
Using Regression with One Predictor .....	99
Comparing Averages for One Variable .....	103
Comparing Proportions .....	105

	Comparing Averages for Multiple Variables .....	108
	Using Regression with Multiple Predictors .....	112
<b>5</b>	<b>Saving and Sharing Your Work</b>	
	<b>Saving and Recreating Results</b> .....	119
	Saving Platform Results in Journals .....	121
	Example: Creating a Journal .....	121
	Adding Additional Analyses .....	122
	Creating Projects .....	122
	Using Scripts .....	123
	Example: Saving and Running a Script .....	123
	About Scripts and JSL .....	125
	Creating Adobe Flash Versions of the Profiler and Bubble Plot .....	125
	Example: Saving an Adobe Flash Version of a Bubble Plot .....	125
<b>6</b>	<b>Special Features</b>	
	<b>Automatic Updating and Integrating with SAS</b> .....	129
	Automatically Updating Analyses and Graphs .....	131
	Example: Using Automatic Recalc .....	131
	Changing Preferences .....	134
	Example: Changing Preferences .....	135
	Integrating JMP and SAS .....	137
	Example: Creating SAS Code .....	138
	Example: Submitting SAS Code .....	138
	<b>Index</b>	
	<b>Discovering JMP</b> .....	141



# 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, Melinda Thielbar, Clay Barker, Peng Liu, Dave Barbour, Jeff Polzin, John Ponte, and Steve Amerige.

In the SAS Institute Technical Support division, Duane Hayes, Wendy Murphrey, Rosemary Lucas, Win LeDinh, Bobby Riggs, Glen Grimme, Sue Walsh, Mike Stockstill, Kathleen Kiernan, and Liz Edwards provide technical support.

Nicole Jones, Kyoko Keener, Hui Di, Joseph Morgan, Wenjun Bao, Fang Chen, Susan Shao, Michael Crotty, Jong-Seok Lee, Tonya Mauldin, Audrey Ventura, Ani Eloyan, Bo Meng, and Sequola McNeill provide ongoing quality assurance. Additional testing and technical support are provided by Noriki Inoue, Kyoko Takenaka, Yusuke Ono, Masakazu Okada, and Naohiro Masukawa from SAS Japan.

Bob Hickey and Jim Borek are the release engineers.

The JMP books were written by Ann Lehman, Lee Creighton, John Sall, Bradley Jones, Erin Vang, Melanie Drake, Meredith Blackwelder, Diane Perhac, Jonathan Gatlin, Susan Conaghan, and Sheila Loring, 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 Yuri 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**

Trish O'Grady coordinates 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; François Bergeret for reviewing the French 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; RuiQi Qiao, Rula Li, and Molly Li for reviewing Simplified Chinese translation (SAS R&D Beijing); Finally, thanks to all the members of our outstanding translation and engineering 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, 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**

Scintilla is Copyright 1998-2003 by Neil Hodgson <neilh@scintilla.org>. NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

XRender is Copyright © 2002 Keith Packard. KEITH PACKARD DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL KEITH PACKARD BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

ImageMagick software is Copyright © 1999-2010 ImageMagick Studio LLC, a non-profit organization dedicated to making software imaging solutions freely available.

bzlib software is Copyright © 1991-2009, Thomas G. Lane, Guido Vollbeding. All Rights Reserved.

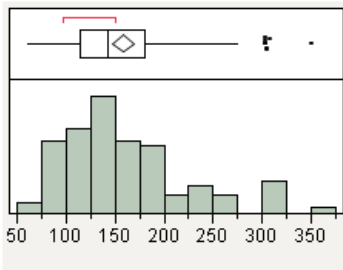
FreeType software is Copyright © 1996-2002 The FreeType Project ([www.freetype.org](http://www.freetype.org)). All rights reserved.



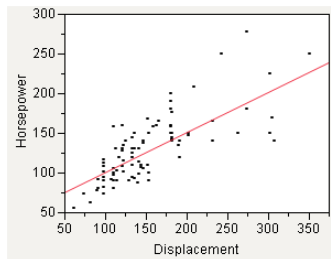
# Gallery of JMP Graphs

## Various Graphs and their Platforms

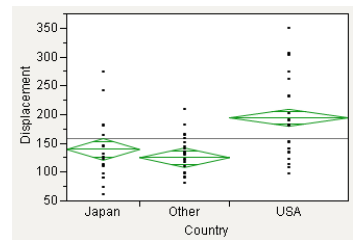
Here are pictures of many of the graphs that you can create with JMP. Each picture is labeled with the platform used to create it. For more information about the platforms and these and other graphs, see the documentation on the **Help > Books** menu.



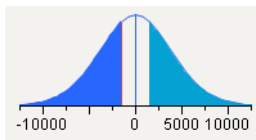
**Histogram**  
Analyze > Distribution



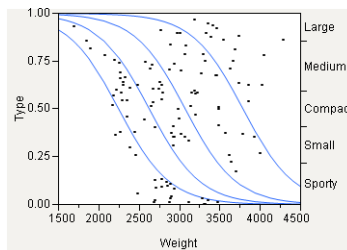
**Bivariate**  
Analyze > Fit Y by X



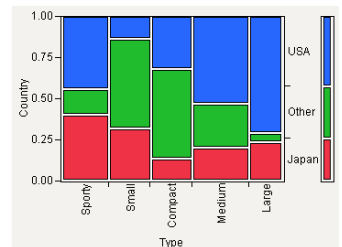
**Oneway**  
Analyze > Fit Y by X



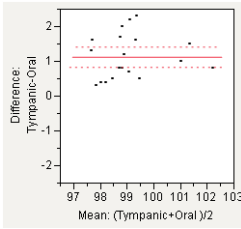
**t Test**  
Analyze > Fit Y by X



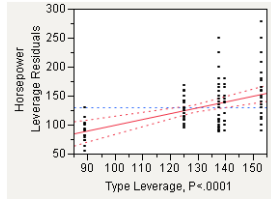
**Logistic**  
Analyze > Fit Y by X



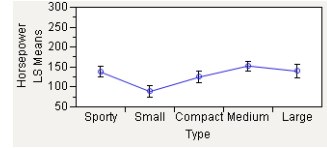
**Mosaic Plot**  
Analyze > Fit Y by X



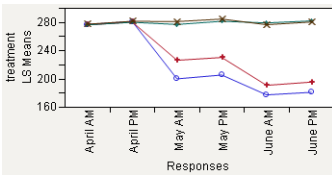
**Matched Pairs**  
Analyze > Matched Pairs



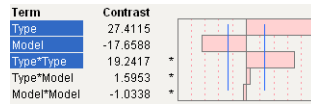
**Leverage Plot**  
Analyze > Fit Model



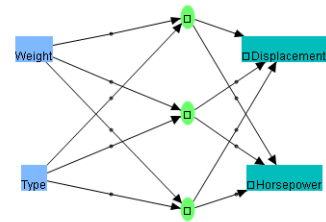
**LS Means Plot**  
Analyze > Fit Model



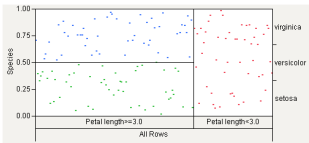
**MANOVA**  
Analyze > Fit Model



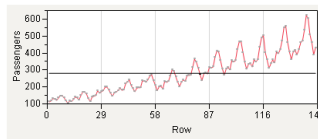
**Screening**  
Analyze > Modeling > Screening



**Neural Diagram**  
Analyze > Modeling > Neural



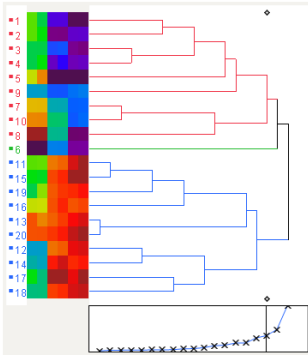
**Partition**  
Analyze > Modeling > Partition



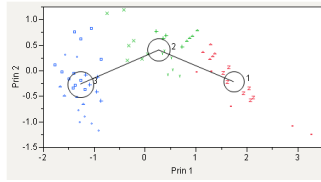
**Time Series**  
Analyze > Modeling > Time Series



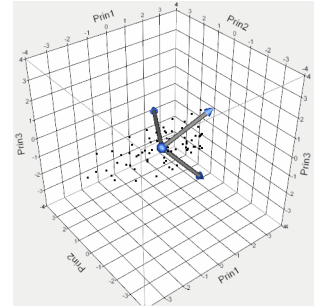
**Share Chart**  
Analyze > Modeling > Categorical



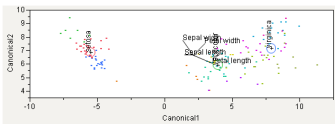
**Dendrogram**  
 Analyze >  
 Multivariate Methods >  
 Cluster



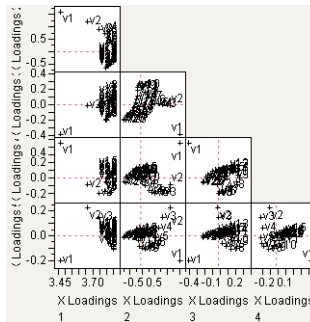
**Self Organizing Map**  
 Analyze >  
 Multivariate Methods >  
 Cluster



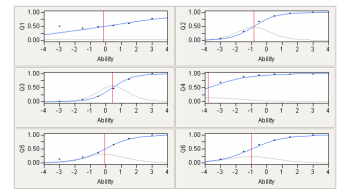
**Principal Components**  
 Analyze >  
 Multivariate Methods >  
 Principal Components



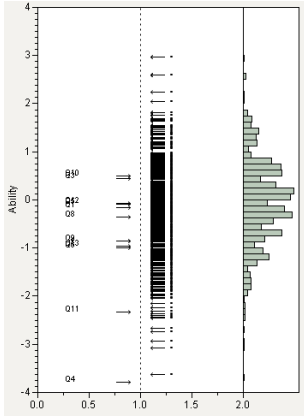
**Canonical Plot**  
 Analyze >  
 Multivariate Methods >  
 Discriminant



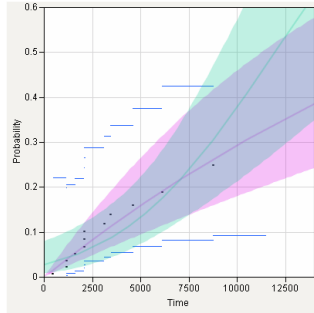
**Loadings Plot**  
 Analyze >  
 Multivariate Methods > PLS



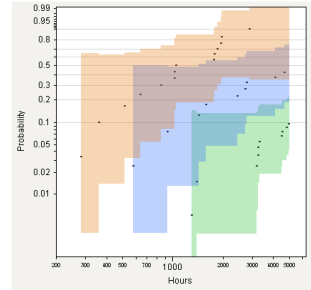
**Characteristic Curves**  
 Analyze >  
 Multivariate Methods > Item  
 Analysis



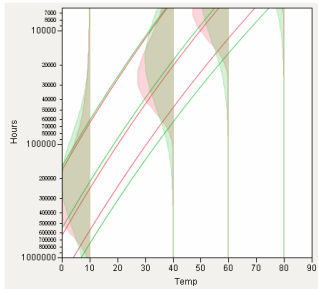
**Dual Plot**  
 Analyze >  
 Multivariate Methods > Item  
 Analysis



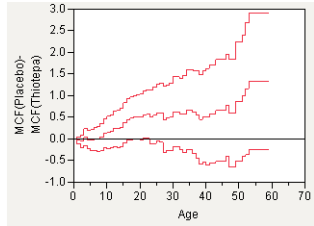
**Compare Distributions**  
 Analyze > Reliability and  
 Survival > Life Distribution



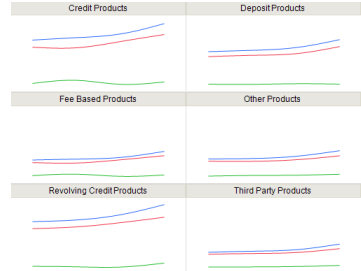
**Nonparametric Overlay**  
 Analyze > Reliability and  
 Survival > Fit Life by X



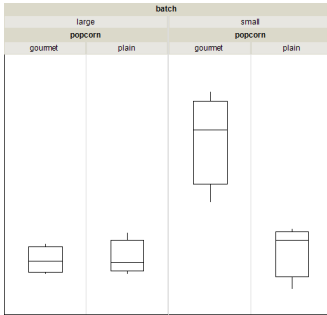
**Scatterplot**  
 Analyze > Reliability and  
 Survival > Fit Life by X



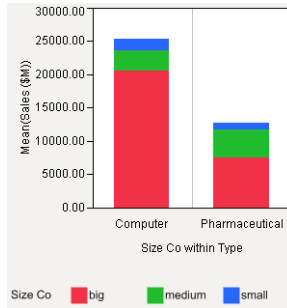
**MCF Plot**  
 Analyze > Reliability and  
 Survival > Recurrence  
 Analysis



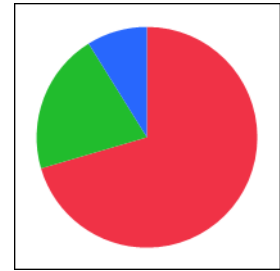
**Line Graphs**  
 Graph > Graph Builder



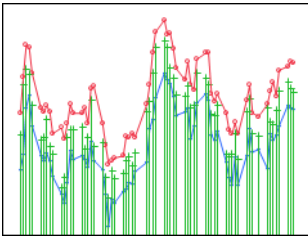
**Box Plots**  
Graph > Graph Builder



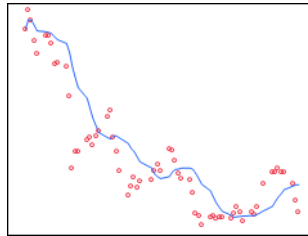
**Stacked Bar Chart**  
Graph > Chart



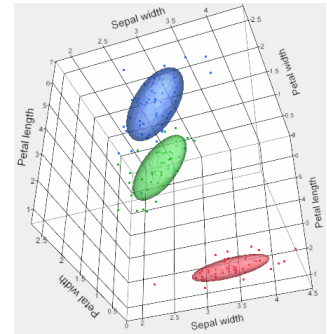
**Pie Chart**  
Graph > Chart



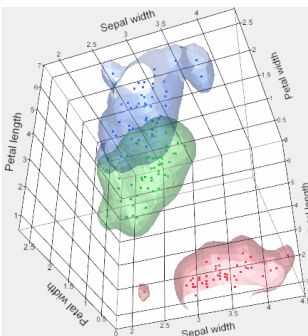
**Needle and Line Chart**  
Graph > Overlay Plot



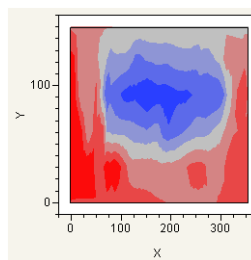
**Dot and Line Chart**  
Graph > Overlay Plot



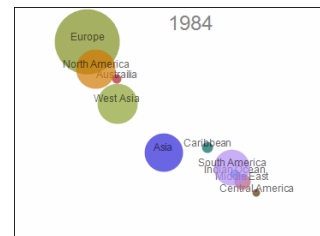
**Three Dimensional Scatterplot**  
Graph > Scatterplot 3D



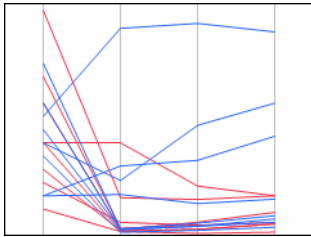
**Three Dimensional Scatterplot**  
Graph > Scatterplot 3D



**Contour Plot**  
Graph > Contour Plot



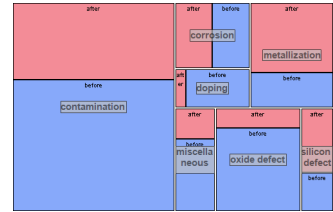
**Bubble Plot**  
Graph > Bubble Plot



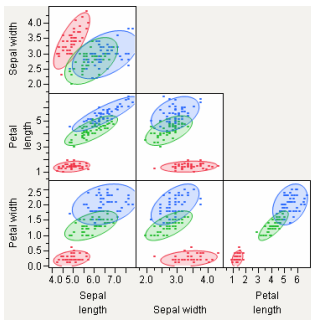
**Parallel Plot**  
Graph > Parallel Plot



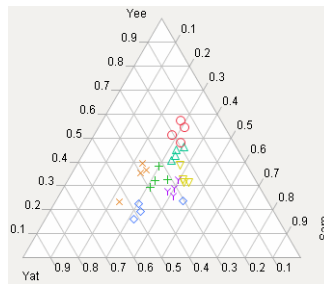
**Cell Plot**  
Graph > Cell Plot



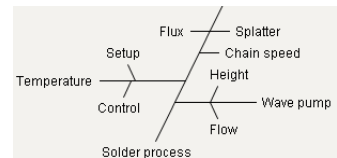
**Tree Map**  
Graph > Tree Map



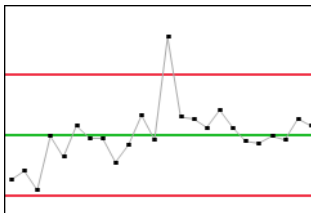
**Scatterplot Matrix**  
Graph > Scatterplot Matrix



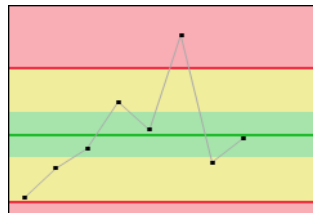
**Ternary Plot**  
Graph > Ternary Plot



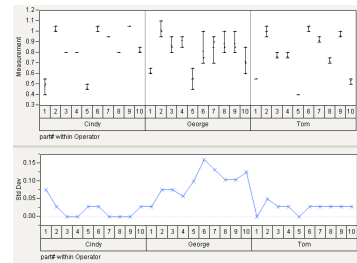
**Ishikawa Chart**  
**Fishbone Chart**  
Graph > Diagram



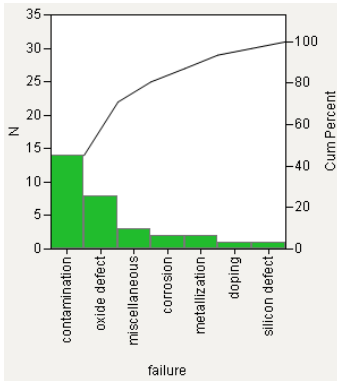
**Individual Measurement Chart**  
**Moving Range Chart**  
Graph > Control Chart > IR



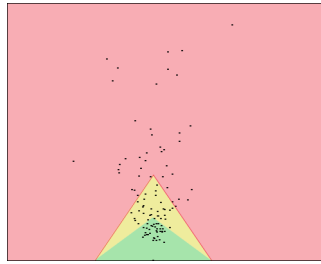
**XBar Chart**  
Graph > Control Chart > XBar



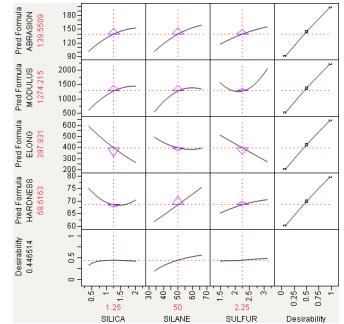
**Variability Chart**  
Graph > Variability/Gauge Chart



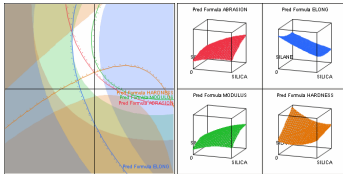
**Pareto Plot**  
Graph > Pareto Plot



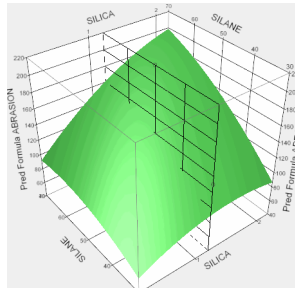
**Goal Plot**  
Graph > Capability



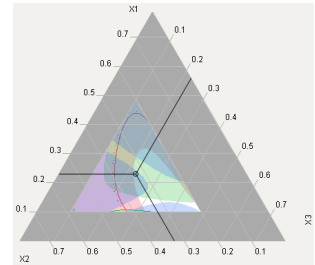
**Prediction Profiler**  
Graph > Profiler



**Contour Profiler**  
Graph > Contour Profiler



**Surface Plot**  
Graph > Surface Plot



**Mixture Profiler**  
Graph > Mixture Profiler



# 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
- 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, *Working with Your 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 multiple-variable graphs that enable you to see relationships between many variables.
- Chapter 4, *Analyzing Your Data*, describes many commonly used analysis techniques. These techniques range from simple techniques that do not require the use of statistical methods, to advanced techniques, where knowledge of statistics is useful.
- Chapter 5, *Saving and Sharing Your Work*, describes using journals and projects, and saving scripts.
- Chapter 6, *Special Features*, describes how to automatically update graphs and analyses as data changes, how to use preferences to customize your reports, 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 for both Windows and Macintosh operating systems, the material in this guide is based on a Windows operating system.



# Chapter 1

## Introducing JMP Basic Concepts

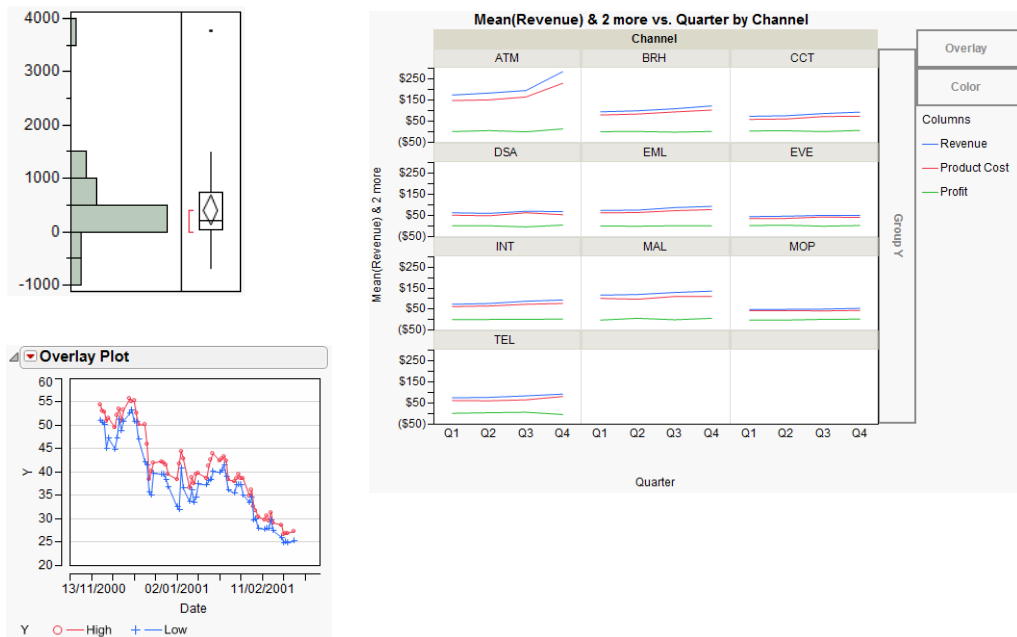
JMP (pronounced jump) is a powerful and interactive data visualization and statistical analysis tool. Use JMP to learn more about your data by performing analyses and interacting with the data using data tables, graphs, charts, and reports.

JMP is useful to the researcher who wants to perform a wide range of statistical analyses and modeling. JMP is equally useful to 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 do the following:

- 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 to predict the future.

**Figure 1.1** Examples of JMP Reports





# Contents

Concepts You Should Know . . . . .	13
How Do I Get Started? . . . . .	13
Starting JMP . . . . .	14
Using Sample Data . . . . .	16
Understanding Data Tables . . . . .	16
Understanding the JMP Workflow . . . . .	18
Step 1: Launching a Platform and Viewing Results . . . . .	19
Step 2: Removing the Box Plot . . . . .	20
Step 3: Requesting Additional Output. . . . .	21
Step 4: Interacting with Platform Results. . . . .	21
How is JMP Different from Excel? . . . . .	23
Formulas . . . . .	23
Columns Names . . . . .	23
Tables and Worksheets. . . . .	23
The Data Grid. . . . .	23
Analysis and Graph Reports . . . . .	23
How Do I Find Help? . . . . .	23

---

## Concepts You Should Know

Before you begin using JMP, you should be familiar with these concepts:

- Enter, view, edit, and manipulate data using JMP *data tables*.
- Select a *platform* from the **Analyze** and **Graph** menus. Platforms contain interactive windows that you use to analyze data and work with graphs.
- Platforms use these windows:
  - *Launch windows* where you set up and run your analysis.
  - *Report windows* showing the output of your analysis.
- Report windows normally contain the following items:
  - A graph of some type (such as a scatterplot or a chart).
  - Specific *reports* that you can show or hide using the *disclosure button* .
  - Platform *options* that are located within *red triangle menus* .

---

## How Do I Get Started?

The general workflow in JMP is simple:

1. Get your data into JMP.
2. Select a platform and complete its launch window.
3. Explore your results and discover where your data takes you.

This workflow is described in more detail in “Understanding the JMP Workflow,” p. 18.

Typically, you start your work in JMP by using graphs to visualize individual variables and relationships among your variables. Graphs make it easy to see this information, and to see the deeper questions to ask. Then you use analysis platforms to dig deeper into your problems and find solutions.

- The “Working with Your Data” chapter shows you how to get data into JMP.
- The “Visualizing Your Data” chapter shows you how to use some of the useful graphs JMP provides to look more closely at your data.
- The “Analyzing Your Data” chapter shows you how to use some of the analysis platforms.

Each chapter teaches through examples. The following sections in this chapter describe data tables and general concepts for working in JMP.

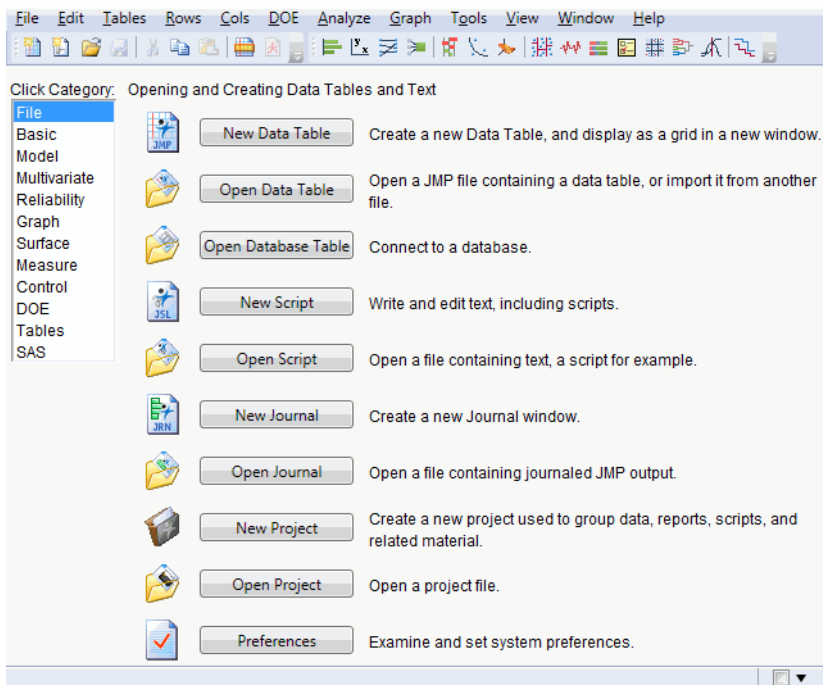
## Starting JMP

Start JMP in two ways:

- Double-click on the JMP icon, normally 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.

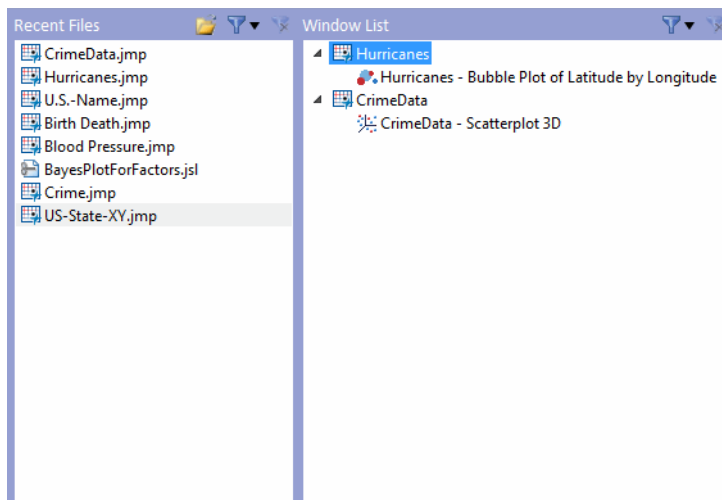
The initial view of JMP includes the Tip of the Day window and the JMP Starter window. The JMP Starter window classifies actions and platforms using categories.

**Figure 1.2** The JMP Starter



On the left is a list of categories. Click a category to see the features and the commands related to that category. For a description of all of the features in the JMP Starter, see *Using JMP*.

Another useful window is the Home window.

**Figure 1.3** The Home Window

To open the Home window, select **View > Home Window**. This window includes links to the following:

- the data tables and report windows that are currently open
- files that you have opened recently

For more details about the JMP Starter window and the Home window, see *Using JMP*.

Almost all JMP windows contain a menu bar and a toolbar. You can find most JMP features in three ways:

- using the menu bar
- using the toolbar buttons
- using the buttons on the JMP Starter window

---

**Note:** By default, windows in JMP are not maximized. This enables you to see the interaction between the windows.

---

### About the Menu Bar and Toolbars

The menus and toolbars are hidden in many windows. To see them, hover your mouse cursor over the blue bar under the window's title bar. The menus in the JMP Starter window, the Home window, and all data tables are always visible.

## Using Sample Data

The examples in this book and the other JMP books use sample data tables. The default location on Windows for the sample data is here:

C:\Program Files\SAS\JMP\9\Support Files English\Sample Data

The Sample Data Index groups the data tables by category. Click a disclosure button to see a list of data tables for that category, and then click a link to open a data table.

### Opening a JMP sample data table

1. From the **Help** menu, select **Sample Data**.
2. Open the **Data Tables for Discovering JMP** list by clicking on the disclosure button next to it.
3. Click the name of the data table to use it in the examples in this book.

### Sample Import Data

Use files from other applications to learn how to import data into JMP.

The default location on Windows for the sample import data is here:

C:\Program Files\SAS\JMP\9\Support Files English\Sample Import Data

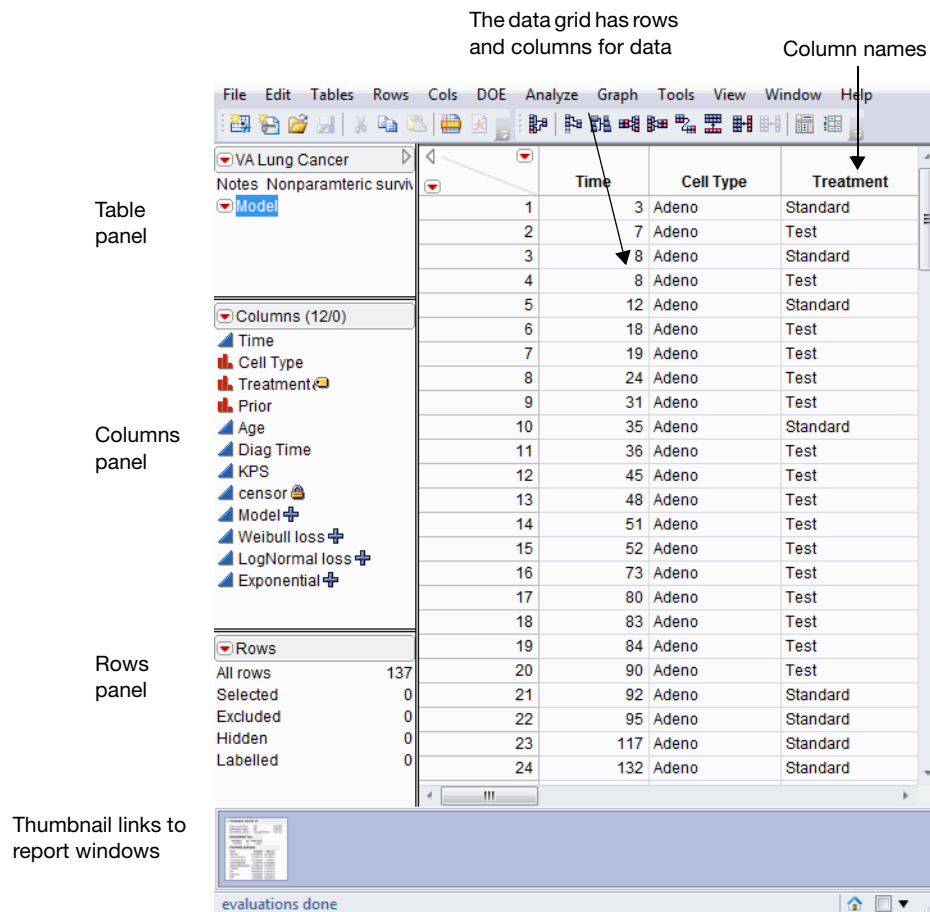
---

## Understanding Data Tables

A data table is a collection of data organized in rows and columns. It is similar to a Microsoft® Excel® spreadsheet, but with some important differences that are discussed in “How is JMP Different from Excel?,” p. 23. A data table might also contain other information like notes, variables, and scripts. These supplementary items are discussed in later chapters.

Open the VA Lung Cancer data table to see the data table described here.

Figure 1.4 A Data Table



A data table contains the following parts:

**Data grid** The data grid contains the data arranged in rows and columns. Generally, each row in the data grid is an observation, and the columns (also called variables) give information about the observations. In Figure 1.4, each row corresponds to a test subject, and there are twelve columns of information. Although all twelve columns cannot be shown in the data grid, the Columns panel lists them all. The information given about each test subject includes the time, cell type, treatment, and more. Each column has a header, or name. That name is not part of the table's total count of rows.

**Table panel** The table panel can contain table variables or table scripts. In Figure 1.4, there is one saved script called **Model** that can automatically recreate an analysis. This table also has a variable

named Notes that contains information about the data. Table variables and table scripts are discussed in a later chapter.

**Columns panel** The columns panel shows the total number of columns, whether any columns are selected, and a list of all the columns by name. The numbers in parentheses (12/0) show that there are twelve columns, and that no columns are selected. An icon to the left of each column name shows that column's modeling type. Modeling types are described in "Understanding Modeling Types," p. 89 in the "Analyzing Your Data" chapter. Icons to the right show any attributes assigned to the column. See "Viewing or Changing Column Information," p. 37 in the "Working with Your Data" chapter for more information about these icons.

**Rows panel** The rows panel shows the number of rows in the data table, and how many rows are selected, excluded, hidden, or labeled. In Figure 1.4, there are 137 rows in the data table.

**Thumbnail links to report windows** This area shows thumbnails of all reports based on the data table. Hover over one to see a larger preview of the report window. Double-click a thumbnail to bring the report window to the front.

Interacting with the data grid, which includes adding rows and columns, entering data, and editing data, is discussed in the "Working with Your Data" chapter. If you open multiple data tables, each one appears in a separate window.

---

## Understanding the JMP Workflow

Once your data is in a data table, you can create graphs or plots, and perform analyses. All features are located in platforms, which are found primarily on the **Analyze** or **Graph** menus. They are called platforms because they do not just produce simple static results. Platform results appear in report windows, are highly interactive, and are linked to the data table and to each other.

The platforms under the **Analyze** and **Graph** menus provide a variety of analytical features and data exploration tools.

The general steps to produce a graph or analysis are as follows:

1. Open a data table.
2. Select a platform from the Graph or Analysis menu.
3. Complete the platform launch window to set up your analysis.
4. Click OK to create the report window that contains your graphs and statistical analyses.
5. Customize your report by using report options.
6. Save, export, and share your results with others.

Later chapters discuss these concepts in greater detail.

The following example shows you how to perform a simple analysis and customize it in four steps. This example uses the `Companies.jmp` file sample data table to show a basic analysis of the variable `Profits ($M)`.

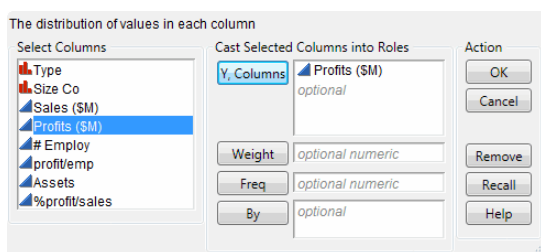
## Step 1: Launching a Platform and Viewing Results

1. Open the Companies.jmp data table.
2. Select **Analyze > Distribution** to open the Distribution launch window.
3. Select Profits (\$M) in the Select Columns box and click the **Y, Columns** button.

The variable Profits (\$M) appears in the **Y, Columns** role. See Figure 1.5 for the completed window.

Another way to assign variables is to click and drag columns from the Select Columns box to any of the roles boxes.

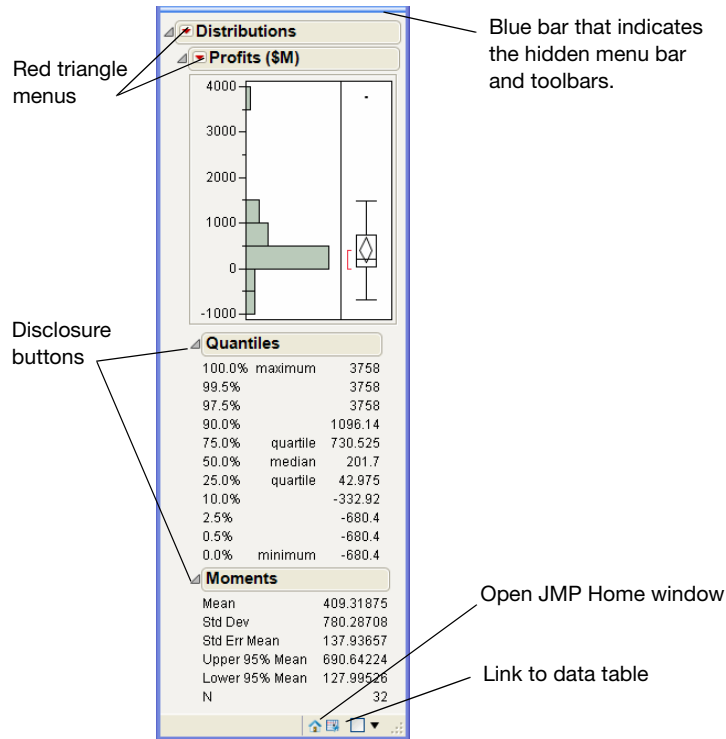
**Figure 1.5** Assign Profits (\$M)



4. Click **OK**.

The Distribution report window appears.

**Figure 1.6** Distribution Report Window



The report window contains basic plots or graphs and preliminary analysis reports. The results appear in an outline format, and you can show or hide any report by clicking on the disclosure button.

Red triangle menus contain options and commands to request additional graphs and analyses at any time.

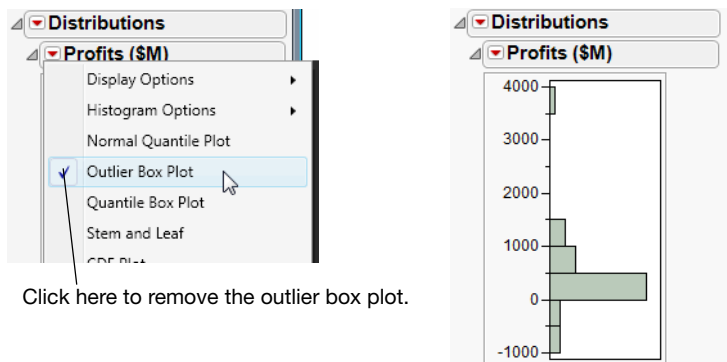
- Hover over the blue bar at the top of the window to see the menu bar and the toolbars.
- Click the data table button to bring the data table that was used to create this report to the front.
- Click the Home window button to return to the Home window.

## Step 2: Removing the Box Plot

Continue using the Distribution report that you created earlier.

1. Click the red triangle next to Profits (\$M) to see a menu of report options.
2. Deselect **Outlier Box Plot** to turn the option off.

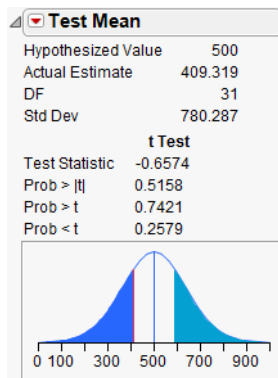
The outlier box plot is removed from the report window.

**Figure 1.7** Removing the Outlier Box Plot

### Step 3: Requesting Additional Output

Continue to use the same report window.

1. From the red triangle menu next to Profits (\$M), select **Test Mean**.  
The Test Mean window appears.
2. Enter 500 in the **Specify Hypothesized Mean** box.
3. Click **OK**.  
The test for the mean is added to the report window.

**Figure 1.8** Test for the Mean

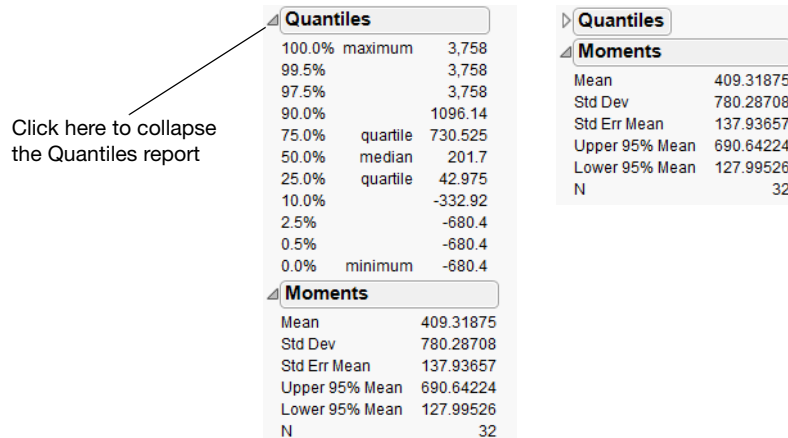
### Step 4: Interacting with Platform Results

All platforms produce results that are interactive. For example:

- Reports can be shown or hidden.
- Additional graphs and statistical details can be added or removed to suit your purposes.
- Platform results are connected to the data table and to each other.

For example, to close the **Quantiles** report, click the disclosure button next to **Quantiles**.

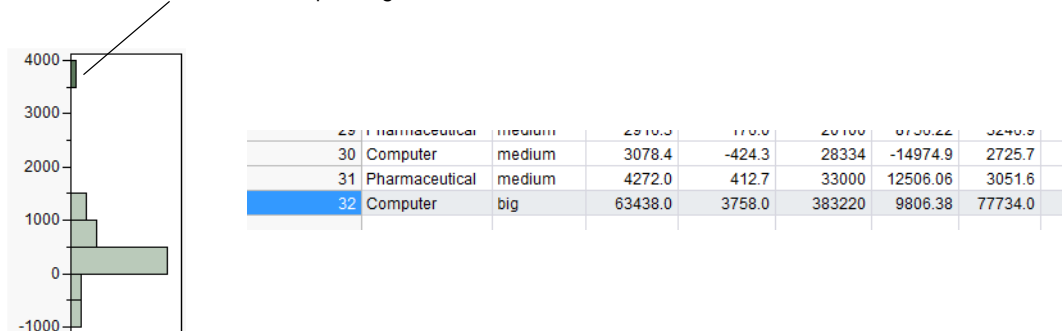
**Figure 1.9** Close the Quantiles Report



Platform results are connected to the data table. The histogram in Figure 1.10 shows that a group of companies makes a much higher profit than the others. To quickly identify that group, click on the histogram bar for them. The corresponding rows in the data table are selected.

**Figure 1.10** Connection Between Platform Results and Data Table

Click the bar to select the corresponding rows



In this case, the group includes only one company, and that one row is selected.

---

## How is JMP Different from Excel?

There are a number of important differences between JMP and Excel or other spreadsheet applications.

### Formulas

**Excel** Formulas are applied to individual cells.

**JMP** Formulas are applied only to entire columns. “Calculating Values With Formulas,” p. 39 in the “Working with Your Data” chapter describes how to use formulas.

### Columns Names

**Excel** Column names are part of the grid. Numbered rows and labeled columns extend past the data. Numeric and character data reside in the same column.

**JMP** Column names are not part of the grid. There are no rows and columns beyond the existing data. The grid is only as big as the data. A column is either numeric or character. If a column contains both character and numeric data, the entire column’s data type is character, and the numbers are treated as character data.

“Understanding Modeling Types,” p. 89 in the “Analyzing Your Data” chapter describes how data type influences platform results.

### Tables and Worksheets

**Excel** A single spreadsheet contains several tables, or worksheets.

**JMP** JMP does not have the concept of worksheets. Each data table is a separate .jmp file and appears in a separate window.

### The Data Grid

**Excel** Data can be located anywhere in the data grid.

**JMP** Data always begins in row 1 and column 1.

### Analysis and Graph Reports

**Excel** All data, analyses, and graphs are placed inside the data grid.

**JMP** Results appear in a separate window.

---

## How Do I Find Help?

As you start using JMP, a variety of resources are available to supplement your learning. The JMP books, Help, web information, sample data, and more are all available on the Help menu.

**Table 1.1** Help Menu Options

Option	Description
<b>Contents, Search, and Index</b>	These three menu items open the JMP Help system. The Help system provides navigable and searchable documentation.
<b>Tip of the Day</b>	A collection of helpful tips and hints that enhance your experience with JMP.
<b>Tutorials</b>	Interactive tutorials that demonstrate how to use some of JMP's statistical and graphical features.  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.
<b>Books</b>	Contains links to the full documentation book set, a description of all the JMP menus, and a quick reference describing keyboard shortcuts.
<b>Sample Data</b>	Provides links to all the sample data used in the documentation. The sample data tables help when you are learning JMP.
<b>Statistics Index</b>	Provides definitions of statistical terms.
<b>JSL Functions</b>	Describes JSL functions and provides examples and topic help.
<b>Object Scripting</b>	Describes JSL objects and messages and provides examples and topic help.
<b>DisplayBox Scripting</b>	Describes JSL display box objects.
<b>JMP.com</b>	Takes you to the JMP Web site ( <a href="http://www.jmp.com">www.jmp.com</a> ). The Web site contains information about JMP 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 find podcasts and other JMP literature. The Web site also contains videos that demonstrate the newest JMP features. You can also learn how to join one of the regional JMP User's Groups.
<b>JMP User Community</b>	Takes you to the online user forum where you can download JMP files, read the JMP Blog, and discuss JMP topics with other users.
<b>About JMP</b>	Shows the release, copyright, operating system, and the owner of the copy of JMP that is running.

# Chapter 2

## Working with Your Data

### Preparing Your Data for Graphing and Analyzing

Before graphing or analyzing your data, the data has to be in a data table and in the proper format. This chapter shows some basic data management tasks, including the following:

- creating new data tables
- opening existing data tables
- importing data from other applications into JMP
- managing your data

**Figure 2.1** Example of a Data Table

	Type	Size Co	Sales (\$M)	Profits (\$M)	# Employ	profit/emp	Assets	%profit/sales
1	Computer	small	855.1	31.0	7523	4120.70	615.2	
2	Pharmaceutical	big	5453.5	859.8	40929	21007.11	4851.6	
3	Computer	small	2153.7	153.0	8200	18658.54	2233.7	
4	Pharmaceutical	big	6747.0	1102.2	50816	21690.02	5681.5	
5	Computer	small	5284.0	454.0	12068	37620.15	2743.9	
6	Pharmaceutical	big	9422.0	747.0	54100	13807.76	8497.0	
7	Computer	small	2876.1	333.3	9500	35084.21	2090.4	
8	Computer	small	709.3	41.4	5000	8280.00	468.1	
9	Computer	small	2952.1	-680.4	18000	-37800.0	1860.7	
10	Computer	small	784.7	89.0	4708	18903.99	955.8	
11	Computer	small	1324.3	-119.7	13740	-8711.79	1040.2	
12	Pharmaceutical	medium	4175.6	939.5	28200	33315.60	5848.0	
13	Computer	big	11899.0	829.0	95000	8726.32	10075.0	
14	Computer	small	873.6	79.5	8200	9695.12	808.0	
15	Pharmaceutical	big	9844.0	1082.0	83100	13020.46	7919.0	
16	Pharmaceutical	small	969.2	227.4	3418	66530.13	784.0	
17	Pharmaceutical	medium	6698.4	1495.4	34400	43470.93	6756.7	
18	Computer	big	5956.0	412.0	56000	7357.14	4500.0	
19	Pharmaceutical	big	5903.7	681.1	42100	16178.15	8324.8	
20	Computer	medium	2959.3	252.8	31404	8049.93	5611.1	
21	Pharmaceutical	small	1198.3	86.5	8527	10144.25	1791.7	
22	Computer	small	800.5	20.0	8578	2426.47	624.2	

# Contents

Getting Your Data Into JMP . . . . .	27
Copying and Pasting Data . . . . .	27
Importing Data . . . . .	27
Entering Data . . . . .	29
Working with Data Tables . . . . .	32
Editing Data . . . . .	32
Selecting, Deselecting, and Finding Values . . . . .	33
Viewing or Changing Column Information . . . . .	37
Calculating Values With Formulas . . . . .	39
Filtering Data . . . . .	41
Managing Data . . . . .	43
Requesting Summary Statistics . . . . .	43
Creating Subsets . . . . .	47
Joining Data Tables . . . . .	49
Sorting Tables . . . . .	51

---

## Getting Your Data Into JMP

- To copy and paste data from another application, see “Copying and Pasting Data,” p. 27.
- To import data from another application, see “Importing Data,” p. 27.
- To enter data directly into a data table, see “Entering Data,” p. 29
- To open a data table, double-click on the file, or use the **File > Open** command.

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

This chapter uses sample data tables and sample import data that is installed with JMP. To find these files, see “Using Sample Data,” p. 16 in the “Introducing JMP” chapter.

### Copying and Pasting Data

You can move data into JMP by copying and pasting from another application, such as Excel or a text file.

#### Example: Copying and pasting data

1. Open the VA Lung Cancer.xls file in Excel. This file is located in the Sample Import Data folder.
2. Select all of the rows and columns, including the column names. There are 12 columns and 138 rows.
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.

If the data that you are pasting into JMP does *not* have column names, then you can use **Edit > Paste**.

### Importing Data

You can move data into JMP by importing data from another application, such as Excel, SAS, or text files. The basic steps to import data are as follows:

1. Select **File > Open**.
2. Navigate to your file’s location.
3. If your file is not listed in the Open Data File window, select the correct file type from the **Files of type** menu.
4. Click **Open**.

#### Example: Importing a Microsoft Excel File

1. Select **File > Open**.
2. Navigate to the Sample Import Data folder.
3. Select VA Lung Cancer.xls.
4. Click **Open**.

The Open Data File window has 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, choose which worksheets to import. Each worksheet is imported to a separate data table.
- Select which version of Excel you need to import from the **Files of type** menu. If the default setting of **All JMP Files** does not show your Excel file, select the correct version from this menu.

When you import Excel files, JMP predicts whether columns headings exist, and if the column names are on row one. The copy and paste method is recommended for the following situations:

- if the column names are located in a row other than row one
- if the file does not include column names and the data does not start in row one
- if the file contains column names and the data does not start in row two

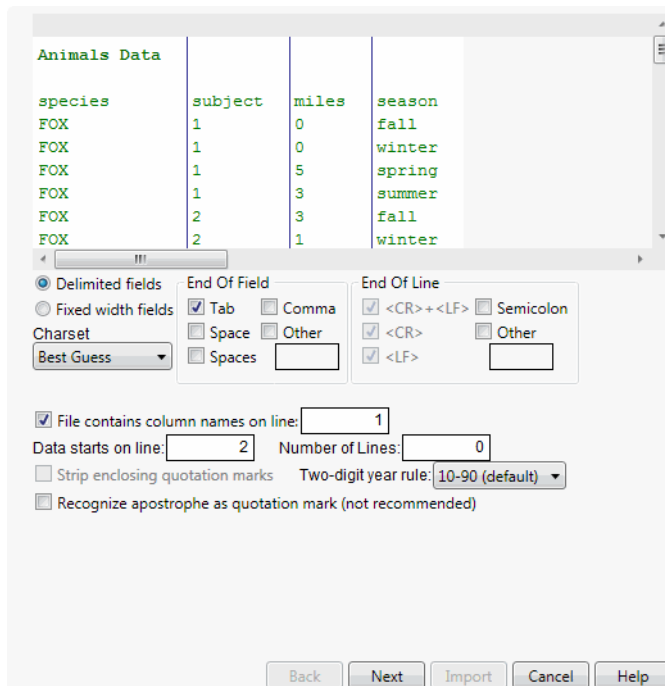
See “Copying and Pasting Data,” p. 27.

#### **Example: Importing a Text File**

One way to import a text file is to let JMP assume the data’s format and place the data in a data table. This method uses settings that you can specify in Preferences. See *Using JMP* for information about setting text import preferences.

Another way to import a text file is to use a Text Preview window to see what your data table will look like after importing, and make adjustments. The following example shows you how to use Text Import Preview window.

1. Select **File > Open**.
2. Navigate to the **Sample Import Data** folder.
3. Select **Animals\_line3.txt**.
4. At the bottom of the Open window, select **Data with preview**.
5. Click **Open**.

**Figure 2.2** Initial Preview Window

This text file has a title on the first line, column names on the third line, and the data starts on line four. If you opened this directly in JMP, the Animals Data line would be the first column name, and all the column names and data afterward would be out of sync. The Preview window lets you adjust the settings before you open the file, and see how your adjustments affect the final data table.

6. Enter 3 in the **File contains column names on line** field.
7. Enter 4 in the **Data starts on line** field.
8. Click **Next**.

In the second window, you can exclude columns from the import and change the data modeling of the columns. For this example, use the default settings.

9. Click **Import**.

The new data table has columns named species, subject, miles, and season. The species and season columns are character data. The subject and miles columns are continuous numeric data.

## Entering Data

You can enter data directly in a data table. The following example shows you how to enter data that was collected over several months into a data table.

## Scenario

Table 2.1 shows the data 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 shows the average blood pressure for each group.

**Table 2.1** Blood Pressure Data

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

## Entering Data in a New Data Table

1. Select **File > New > Data Table** to create an empty data table.

A new data table has one column and no rows.

2. Click the column name to select it, and then click again to edit the name.

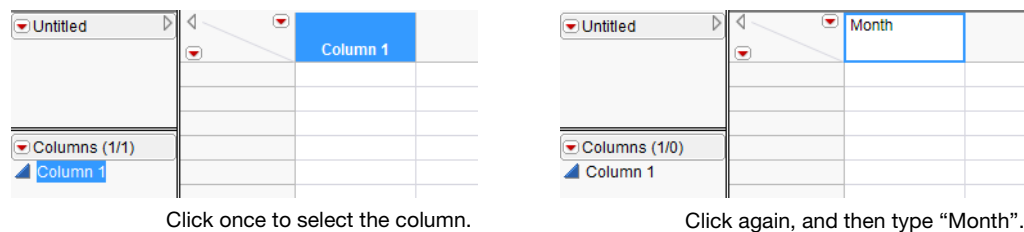
---

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

---

3. Change the column name to **Month**. See Figure 2.3.

**Figure 2.3** Entering a Column Name

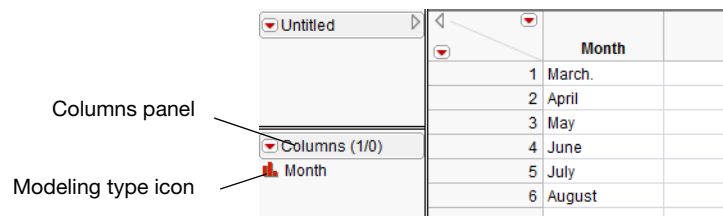


4. Select **Rows > Add Rows**.

The Add Rows window appears.

5. Since you want to add six rows, type 6.
6. Click **OK**. Six empty rows are added to the data table.
7. Enter the Month information by double-clicking in a cell and typing.

**Figure 2.4** Month Column Completed



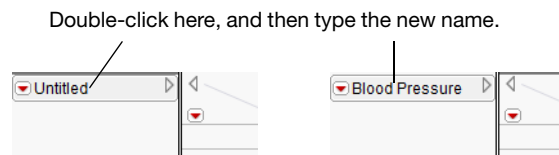
In the columns panel, look at the modeling type icon to the left of the column name. It has changed to reflect that Month is now nominal (previously it was continuous). Compare the modeling type shown for Column 1 in Figure 2.3 and for Month in Figure 2.4. This difference is important and is discussed in “Viewing or Changing Column Information,” p. 37.

8. Double-click in the space on the right side of the Month column to add the Control column.
9. Change the name to Control.
10. Enter the Control data as shown in Table 2.1. Your data table now consists of six rows and two columns.
11. Continue adding columns and entering data as shown in Table 2.1 to create the final data table with six rows and five columns.

### Changing the data table name

1. Double-click on the data table name (Untitled) in the Table Panel.
2. Type the new name (Blood Pressure).

**Figure 2.5** Changing the Data Table Name



---

## Working with Data Tables

This section contains the following information:

- “Editing Data,” p. 32
- “Selecting, Deselecting, and Finding Values,” p. 33
- “Viewing or Changing Column Information,” p. 37
- “Filtering Data,” p. 41
- “Calculating Values With Formulas,” p. 39

### Editing Data

You can enter or change data, either a few cells at a time or for an entire column. This section contains the following information:

- “Changing Values,” p. 32
- “Recoding Values,” p. 32
- “Creating Patterned Data,” p. 33

The examples in this section use the `Companies.jmp` sample data table. Before you continue, open this data table.

### Changing Values

To change a value, select a cell and type the change. You can also double-click a cell to edit it.

---

**Note:** Double-clicking in a cell is not the same as selecting a cell. A single click selects a cell. You can select more than one cell at the same time, and you can perform certain actions on selected cells. Double-clicking only lets you edit a cell. For more information about selecting rows, columns, and cells, see “Selecting, Deselecting, and Finding Values,” p. 33.

---

### Recoding Values

Use the recoding tool to change all of the values in a column at once. For example, suppose you are interested in comparing the sales of computer and pharmaceutical companies. Your current company labels are Computer and Pharmaceutical. You want to change them to Technical and Drug. Going through all 32 rows of data and changing all the values would be tedious, inefficient, and error-prone, especially if you had many more rows of data. Recode is a better option.

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 Technical in the Computer row, and Drug in the Pharmaceutical row.
4. Select the **In Place** option from the menu.

**Figure 2.6** Recode Window

Count	Old Value	New Value
20	Computer	Technical
12	Pharmaceutical	Drug

Buttons: OK, Cancel, In Place (dropdown)

5. Click **OK**.

All cells are updated automatically to the new values.

### Creating Patterned Data

Use the Fill options to populate a column with patterned data. The Fill options are especially useful if your data table is large, and typing in the values for each row would be cumbersome.

#### Example: Filling a column with the pattern

1. Add a new column.
2. Enter 1 in the first cell, 2 in the second cell, and 3 in the third cell.
3. Select the three cells, and right-click anywhere in the selected cells to see a menu.
4. Select **Fill > Repeat sequence to end of table**.

The rest of the column is filled with the sequence (1, 2, 3, 1, 2, 3, ...).

To continue a pattern instead of repeating it (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 options only repeat the values.

### Selecting, Deselecting, and Finding Values

You can select rows, columns, or cells within a data table. For example, to create a subset of an existing data table, you must first select the parts of the table that you want to subset. Also, selecting rows can make data points stand out on a graph. Select rows and columns manually by clicking, or select rows that meet certain search criteria. This section contains the following information:

- “Selecting and Deselecting Rows,” p. 34
- “Selecting and Deselecting Columns,” p. 34
- “Selecting and Deselecting Cells,” p. 35
- “Searching for Values,” p. 36

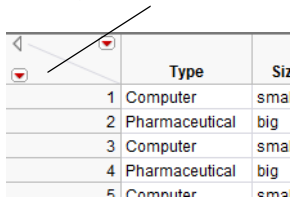
## Selecting and Deselecting Rows

**Table 2.2** Selecting and Deselecting Rows

Task	Action
Select rows one at a time	Click on the row number.
Select multiple adjacent rows	Click and drag on the row numbers.  or Select the beginning row, and then hold down the SHIFT key and click the last row number.
Select multiple non-adjacent rows	Select the first row, and 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 2.7.

**Figure 2.7** Deselecting Rows

To deselect all rows at once, click here.



	Type	Siz
1	Computer	sma
2	Pharmaceutical	big
3	Computer	sma
4	Pharmaceutical	big
5	Computer	ema

## Selecting and Deselecting Columns

**Table 2.3** Selecting and Deselecting Columns

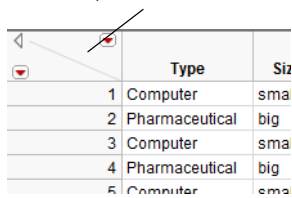
Task	Action
Select columns one at a time	Click the column heading.
Select multiple adjacent columns	Click and drag across the column headings.  or Select the beginning column, and then hold down the SHIFT key and click the last header.

**Table 2.3** Selecting and Deselecting Columns (*Continued*)

Task	Action
Select multiple non-adjacent columns	Select the first column, and 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 2.8.

**Figure 2.8** Deselecting Columns

To deselect all columns at once, click here.



	Type	Size
1	Computer	small
2	Pharmaceutical	big
3	Computer	small
4	Pharmaceutical	big
5	Computer	small

### Selecting and Deselecting Cells

**Table 2.4** Selecting and Deselecting Cells

Task	Action
Select cells one at a time	Click each cell individually.
Select multiple adjacent cells	Click and drag across the cells.  or Select the beginning cell, and then hold down the SHIFT key and click the last cell.
Select multiple non-adjacent cells	Select the first cell, and 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.

## ***Searching for Values***

In a data table that has thousands or tens of thousands of rows, it can be difficult to locate a particular cell by scrolling through the table. If you are looking for specific information, use the Search feature to find it. If data is found that matches the search criteria, the cell is selected and the data grid scrolls to show it in the window. For example, the `Companies.jmp` data table contains information about a company that has total sales of \$11,899. Use the Search feature to find that cell.

### **Example: Searching for a value**

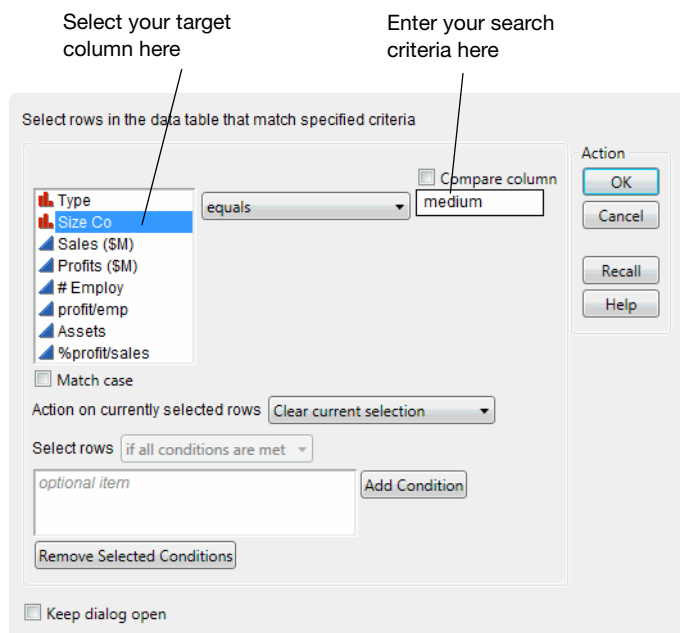
1. Select **Edit > Search > Find** to launch the Search window.
2. In the **Find what** box, enter 11899.
3. Click **Find**. JMP finds the first cell that has 11,899 in it, and selects it.

If multiple cells meet the search criteria, click **Find** again to find the next cell that matches the search term.

You can also search for multiple rows at once, with each row matching some criteria.

### **Example: Select all of the rows that correspond to medium-sized companies**

1. Select **Rows > Row Selection > Select Where** to open 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**.
4. Click **OK**.

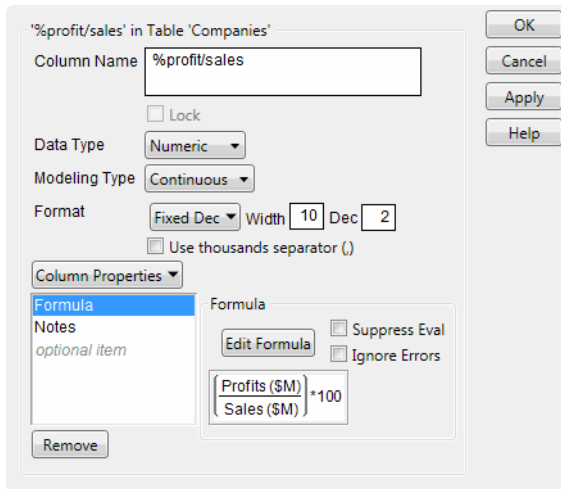
**Figure 2.9** Select Rows Window

JMP selects all of the rows that have Size Co equal to medium. There are seven.

## Viewing or Changing Column Information

Information about a column is not limited to the data in the column. Data type, modeling type, format, and formulas can also be set.

To view or change column characteristics, double-click on the column heading. Or, right-click on the column heading and select **Column Info**. The Column Info window appears.

**Figure 2.10** Column Info Window**Table 2.5** Column Information

Option	Description
<b>Column Name</b>	Enter or change the column name. No two columns can have the same column name.
<b>Data Type</b>	Select one of the following data types: <ul style="list-style-type: none"> <li>• <b>Numeric</b> specifies the column values as numbers.</li> <li>• <b>Character</b> specifies the column values as non-numeric, such as letters or symbols.</li> <li>• <b>Row State</b> specifies the column values as row states. This is an advanced topic. See <i>Using JMP</i>.</li> </ul>
<b>Modeling Type</b>	Modeling types define how values are used in analyses. Select one of the following modeling types: <ul style="list-style-type: none"> <li>• <b>Continuous</b> values are numeric only.</li> <li>• <b>Ordinal</b> values are either numeric or character, and are ordered categories.</li> <li>• <b>Nominal</b> values are either numeric or character, but not ordered.</li> </ul>

**Table 2.5** Column Information (*Continued*)

Option	Description
<b>Format</b>	Select a format for numeric values. This option is not available for character data. Here are a few of the most common formats: <ul style="list-style-type: none"> <li>• <b>Best</b> lets JMP choose the best display format.</li> <li>• <b>Fixed Dec</b> specifies the number of decimal places that appear.</li> <li>• <b>Date</b> specifies the syntax for date values.</li> <li>• <b>Time</b> specifies the syntax for time values.</li> <li>• <b>Currency</b> specifies the type of currency and decimal points that are used for currency values.</li> </ul>
<b>Column Properties</b>	Set special column properties such as formulas, notes, and value orders. See <i>Using JMP</i> .
<b>Lock</b>	Lock a column, so that the values in the column cannot be changed.

## Calculating Values With Formulas

Use the Formula Editor to create columns that contain calculated values.

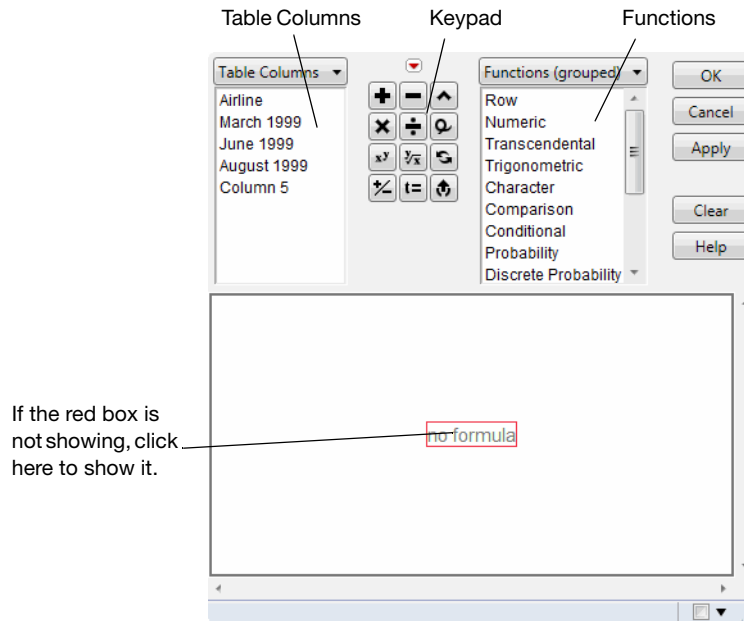
### Scenario

The sample data table *On-Time Arrivals.jmp* reflects the percent of on-time arrivals for several airlines. The data was collected for March, June, and August of 1999.

### Creating the Formula

Suppose that you want to create a new column containing the average on-time percentage for each airline.

1. Add a new column.
2. Right-click on the column heading of the new column and select **Formula**. The Formula Editor window appears.

**Figure 2.11** Formula Editor


Create the formula for the average on-time percentage of each airline:

3. From the Table Columns list, select **March 1999**.
4. Click the **+** button on the keypad.
5. Select **June 1999**, followed by another **+** sign.
6. Select **August 1999**.

**Figure 2.12** Sum of the Months


*March 1999* + *June 1999* + *August 1999*

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

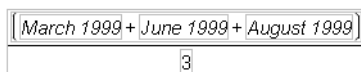
7. Click on the box surrounding the entire formula.

**Figure 2.13** Entire Formula Selected

*March 1999* + *June 1999* + *August 1999*

8. Click the  button.
9. Type a 3 in the denominator box, and then click outside of the formula in any of the white space.

**Figure 2.14** Completed Formula



10. Click **OK**

The new column contains the averages.

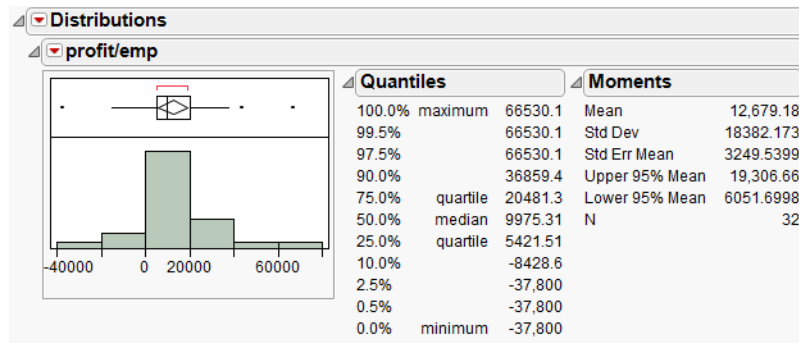
The Formula Editor has many built-in arithmetic and statistical functions. For example, another way to calculate the average on-time arrival percentage is to use the **Mean()** function in the Statistical functions list. For details about all of the Formula Editor functions, see *Using JMP*.

## Filtering Data

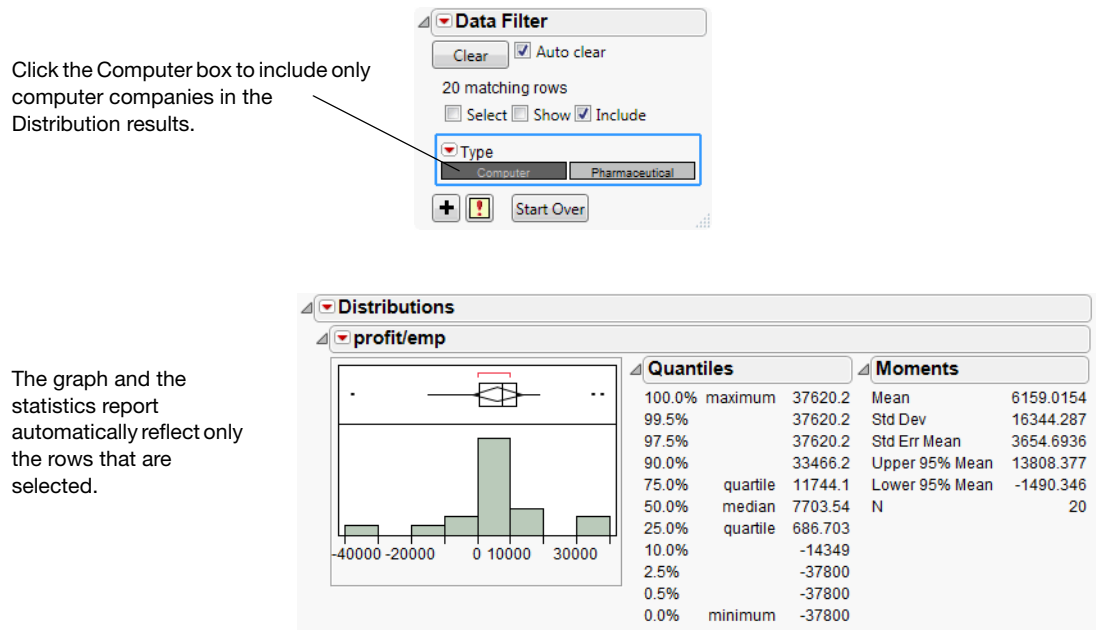
Use the **Data Filter** to interactively select complex subsets of data, hide these subsets in plots, or exclude them from analyses. For example, look at profit per employee for computer and pharmaceutical companies.

1. Open the Companies.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select profit/emp and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu for profit/emp, select **Display Options > Horizontal Layout**.

**Figure 2.15** Distribution of profit/emp



6. Turn on Automatic Recalc by selecting **Script > Automatic Recalc** from the red triangle menu for **Distributions**.  
 When this option is on, every change that you make (for example, hiding or excluding points) causes your report window to automatically update itself.
7. Select **Rows > Data Filter**.
8. Select Type and click **Add**.
9. Clear **Select** and select **Include**.
10. To filter out the Pharmaceutical companies from the Distribution results, and include only the Computer companies, click the **Computer** box in the Data Filter window.  
 The distribution results update to only include Computer companies.

**Figure 2.16** Filter for Computer Companies


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

---

## Managing Data

The commands on the **Tables** menu summarize and manipulate data tables into the format that you need for graphing and analyzing. This section describes five of the commands:

**Summary** Creates a table that contains summary statistics that describe your data.

**Tabulate** Provides a drag and drop workspace to create summary statistics.

**Subset** Creates a table that contains a subset of your data.

**Join** Joins the data from two data tables into one new data table.

**Sort** Sorts your data by one or more columns.

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

## Requesting Summary Statistics

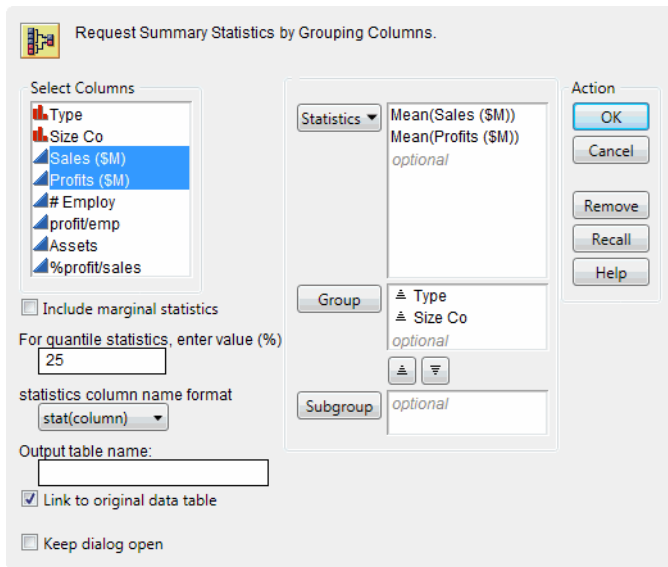
Summary statistics, such as sums and means, can instantly provide useful information about your data. For example, if you look at the annual profit of each company out of thirty-two companies, it's difficult to compare the profits of small, medium, and large companies. A summary shows that information immediately.

Create summary tables by using either the Summary or Tabulate commands on the Tables menu. The Summary command creates a new data table. As with any data table, you can perform analyses and create graphs from the summary table. The Tabulate command creates a report window with a table of summary data. You can also create a table from the Tabulate report.

### Summary

A summary table contains statistics for each level of a grouping variable. For example, look at the financial data for computer and pharmaceutical companies. Suppose that 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. Select `Type` and `Size Co` and click **Group**.
4. Select `Sales ($M)` and `Profits ($M)` and click **Statistics > Mean**.

**Figure 2.17** Completed Summary Window

5. Click **OK**.

JMP calculates the mean of Sales (\$M) and the mean of Profit (\$M) for each combination of Type and Size Co.

**Figure 2.18** Summary Table

	Type	Size Co	N Rows	Mean(Sales (\$M))	Mean(Profits (\$M))
1	Computer	big	4	20,597.48	1,089.93
2	Computer	medium	2	3,018.85	-85.75
3	Computer	small	14	1,758.06	44.94
4	Pharmaceutical	big	5	7,474.04	894.42
5	Pharmaceutical	medium	5	4,261.06	698.98
6	Pharmaceutical	small	2	1,083.75	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 that correspond to each 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 a column for the mean of Profits (\$M).

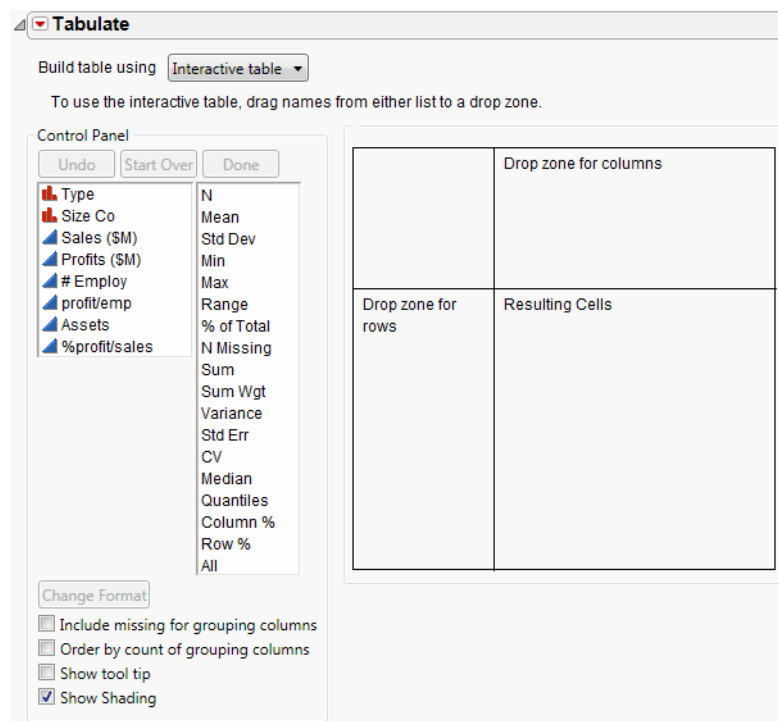
The summary table is linked to the source table. Selecting a row in the summary table also selects the corresponding rows in the source table.

## Tabulate

Use the Tabulate command to drag columns into a workspace, creating summary statistics for each combination of grouping variables. This example shows you how to use Tabulate to create the same summary information that you just created using Summary.

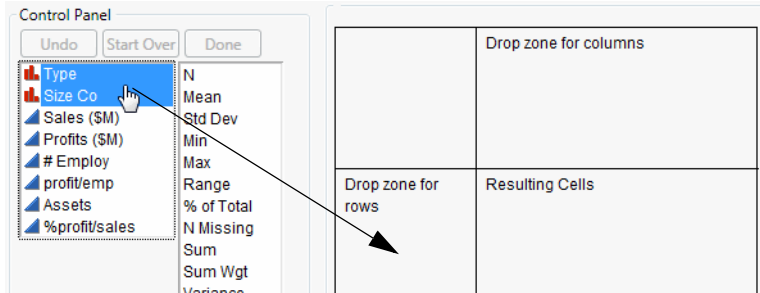
1. Open the Companies.jmp sample data table.
2. Select **Tables > Tabulate**.

**Figure 2.19** Tabulate Workspace



3. Select both Type and Size Co.

4. Drag and drop them into the **Drop zone for rows** to see a menu.

**Figure 2.20** Dragging Columns to the Row Zone


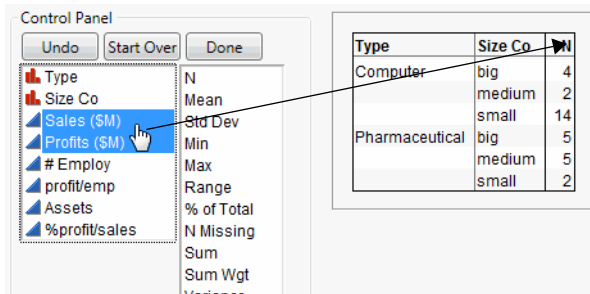
5. Select **Add Grouping Columns**.

The initial tabulation shows the number of rows per group.

**Figure 2.21** Initial Tabulation

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

6. Select both Sales (\$M) and Profits (\$M), and drag and drop them over the **N** in the table to see a menu.

**Figure 2.22** Adding Sales and Profit


7. Select **Add Analysis Columns**.

The tabulation now shows the sum of Sales (\$M) and the sum of Profits (\$M) per group. The final step is to change the sums to means.

**Figure 2.23** Tabulation of Sums

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

8. Right-click on **Sum** (either of them) and select **Statistics > Mean**.

The sums are replaced by the means for each group.

**Figure 2.24** Final Tabulation

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

The means are the same as those obtained using the Summary command. Compare Figure 2.24 to Figure 2.18.

## Creating Subsets

If you want to look closely at only part of your data table, you can create a subset. For example, suppose that you have already compared the sales and profits of big, medium, and small computer and pharmaceutical companies. Now you want to look at the sales and profits of only the medium-sized companies.

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

### ***Subsetting with the Subset Command***

1. Open the Companies.jmp sample data table.

Selecting the rows and columns that you want to subset

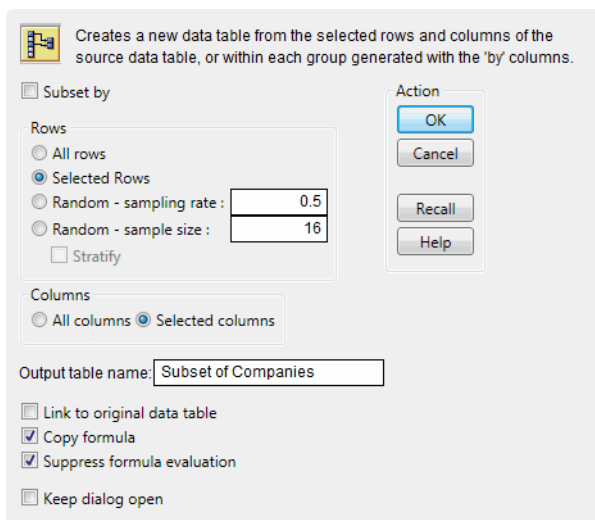
2. Select **Rows > Row Selection > Select Where**.

3. Select Size Co in the column list box on the left.
4. Enter medium in the text enter box.
5. Click **OK**.
6. Select the Type, Sales (\$M), and Profits (\$M) columns.

### Creating the subset table

7. Select **Tables > Subset** to launch the Subset window.

**Figure 2.25** Subset Window



8. Select **Selected columns** to subset only the columns that you selected. You can also customize your subset table further by selecting additional options.
9. Click **OK**.

The resulting subset data table has seven rows and three columns. For complete details about the Subset command, see *Using JMP*.

### **Subsetting with the Distribution Platform**

Another way to create subsets uses the connection between platform results and data tables.

#### **Example: Creating a subset using the Distribution command**

1. Open the Companies.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Type and click **Y, Columns**.

4. Click **OK**.
5. Double-click on the histogram bar that represents Computer to create a subset table of the Computer companies.

---

**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.

---

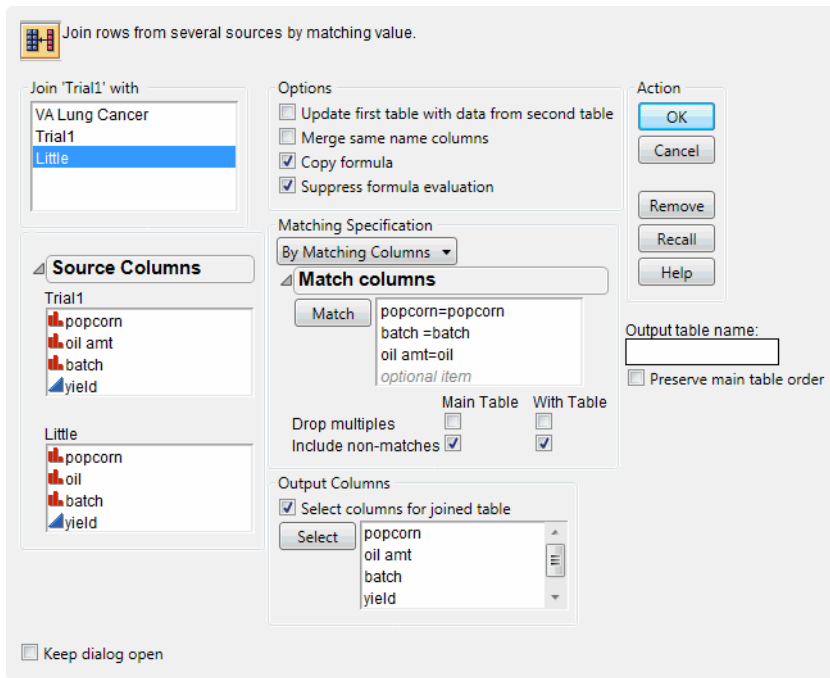
## Joining Data Tables

Use the Join option to combine information from multiple data tables into a single data table. For example, suppose that you have a data table containing results from an experiment on popcorn yields. In another data table, you have the results of a second experiment on popcorn yields. To compare the two experiments or to analyze the trials using both sets of results, you need to have the data in the same table. Also, the experimental data was not entered into the data tables in the same order. One of the columns has a different name, and the second experiment is incomplete. This means that you cannot copy and paste from one table into another.

### Example: Joining two data tables

1. Open the Trial1.jmp and Little.jmp data tables.
2. Click on Trial1.jmp to make it the active data table.
3. Select **Tables > Join**.
4. In the **Join 'Trial1' With** box, select Little.
5. From the **Matching Specification** menu, select **By Matching Columns**.
6. In the **Source Columns** boxes, select popcorn in both boxes, and then click **Match**.
7. In the same way, match batch in both boxes, and then match oil amt for Trial1 and oil for Little.  
Your matching columns do not have to have the same name.
8. Select **Include non-matches** for both tables.  
Since one experiment is partial, you want to include all columns, including any with missing data.
9. To avoid duplicate columns, select the **Select columns for joined table** option.
10. From Trial1, select all four columns and click **Select**.
11. From Little, select only yield and click **Select**.

Figure 2.26 Completed Join Window



12. Click **OK**.

Figure 2.27 Joined Table

	oil amt	batch	yield of Trial1	yield of Little
1	little	large	8.6	8.2
2	lots	large	9.2	•
3	little	small	12.1	15.9
4	lots	small	18.0	•
5	little	large	8.2	8.8
6	lots	large	10.4	•
7	little	small	9.9	10.1
8	lots	small	10.6	•

Columns (4/0)

- oil amt
- batch
- yield of Trial1
- yield of Little

Rows

- All rows: 8
- Selected: 0
- Excluded: 0
- Hidden: 0
- Labelled: 0

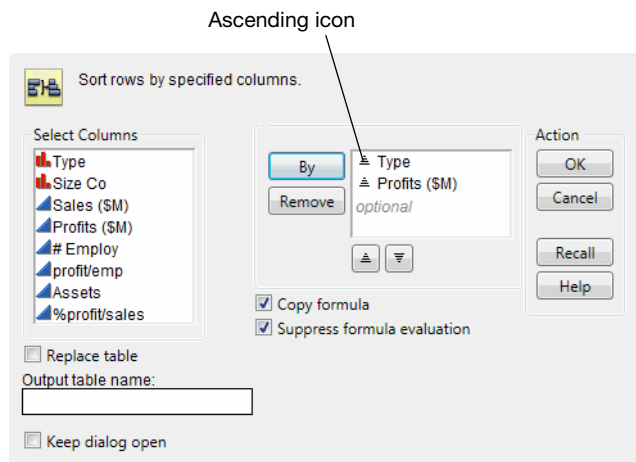
## Sorting Tables

Use the Sort command to sort a data table by one or more columns in the data table. For example, look at financial data for computer and pharmaceutical companies. Suppose that you want to sort the data table by Type, then by Profits (\$M). Additionally, you want Profits (\$M) to be in descending order within each Type.

1. Open the Companies.jmp sample data table.
2. Select **Tables > Sort**.
3. Select Type and click **By** to assign Type as a sorting variable.
4. Select Profits (\$M) and click **By**.

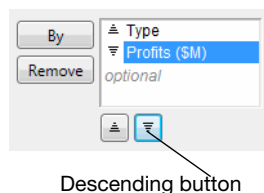
At this point, both variables are set to be sorted in ascending order. See the ascending icon next to the variables in Figure 2.28.

**Figure 2.28** Sort Ascending Icon



5. To change Profits (\$M) to sort in descending order, select Profits (\$M) and click the descending button.

**Figure 2.29** Change Profits to Descending



The icon next to Profits (\$M) changes to descending.

6. Select the **Replace Table** check box.

When selected, the **Replace Table** 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. Sorting a data table with open report windows might change how some of the data is displayed in the report window, especially in graphs.

7. Click **OK**.

The data table is now sorted by type alphabetically, and by descending profit totals within type.

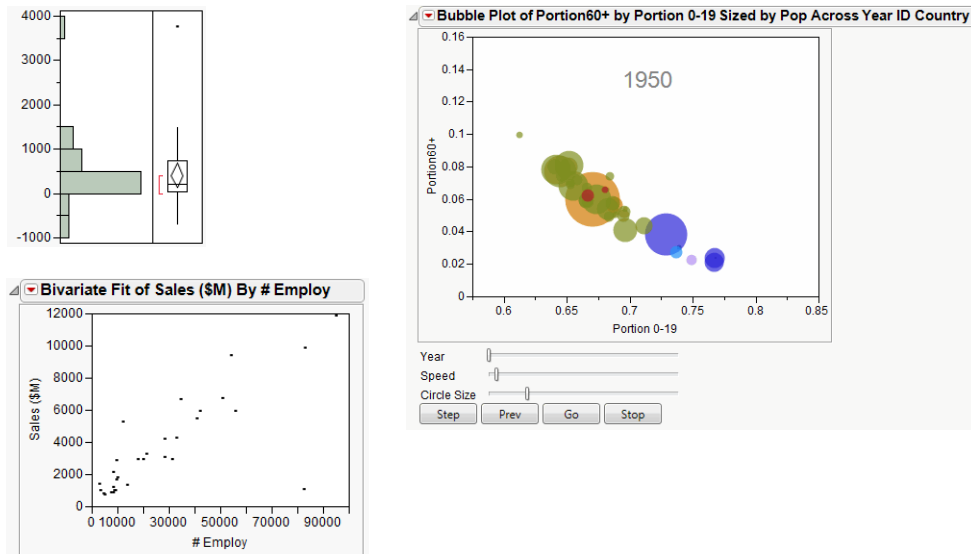
# Chapter 3

## Visualizing Your Data Graphing Data

Visualizing your data is an important first step. The graphs described in this chapter help you to discover important details about your data. For example, histograms show you the shape and range of your data, and help you find unusual data points.

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 some of JMP's graphical tools and platforms. Use JMP to visualize the distribution of single variables, or the relationships among multiple variables.

**Figure 3.1** Visualizing Data with JMP



# Contents

Looking at Single Variables . . . . .	55
Histograms . . . . .	55
Bar Charts . . . . .	57
Comparing Multiple Variables. . . . .	59
Scatterplots . . . . .	60
Scatterplot Matrix . . . . .	64
Side-by-Side Box Plots. . . . .	67
Overlay Plots . . . . .	69
Variability Chart . . . . .	73
Graph Builder . . . . .	75
Bubble Plots . . . . .	80

---

## Looking at Single Variables

Single-variable graphs, or *univariate* graphs, let you look closely at one variable at a time. When you begin to look at your data, it's important to learn about each variable before looking at how the variables interact with each other. Univariate graphs let you visualize each variable individually.

This section covers two graphs that show the distribution of a single variable:

- “Histograms,” p. 55, for continuous variables
- “Bar Charts,” p. 57, for categorical variables

Use the Distribution platform to create both of these graphs. Distribution produces a graphical description and descriptive statistics for each variable.

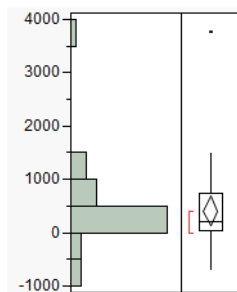
### Histograms

The histogram is one of the most useful graphical tools for understanding the distribution of a continuous variable. Use a histogram to find the following in your data:

- the average value and variation
- extreme values

---

**Figure 3.2** Example of a Histogram



---

### Scenario

This example uses the `Companies.jmp` data table, which contains data on profits for a group of companies.

A financial analyst wants to explore the following questions:

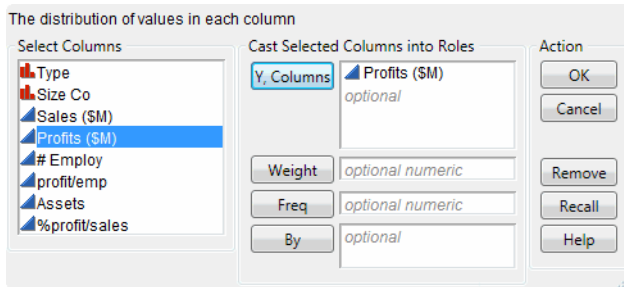
- Generally, how much profit does each company earn?
- What is the average profit?
- Are there any companies that earn either extremely high or extremely low profits compared to the other companies?

To answer these questions, use a histogram of Profits (\$M).

### Creating the Histogram

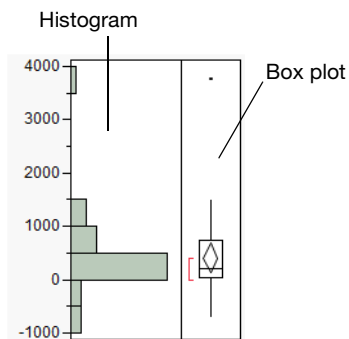
1. Open the Companies.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Profits (\$M) and click **Y, Columns**.

**Figure 3.3** Distribution Window for Profits (\$M)



4. Click **OK**.

**Figure 3.4** Histogram of Profits (\$M)



### Interpreting the Histogram

The histogram provides these answers:

- Most companies' profits are between \$-1000 and \$1500.  
 All the bars except for one are located in this range. Additionally, more companies' profits range from \$0 to \$500 than any other range. The bar representing that range is much longer than the others.

- The average profit is a little less than \$500.  
The middle of the diamond in the box plot indicates the mean value. In this case, the mean is slightly lower than the \$500 mark.
- 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 data points.  
This outlier is represented by a single, very short bar at the top of the histogram. The bar is small and represents a small group (in this case, a single company), and it is widely separated from the rest of the histogram bars.

In addition to the histogram, this report includes the following:

- The box plot, which is another graphical summary of the data. For detailed information about the box plot, see *Basic Analysis and Graphing*.
- **Quantiles** and **Moments** reports. These reports are discussed in “Analyzing Distributions,” p. 93 in the “Analyzing Your Data” chapter.

### Interacting with the Histogram

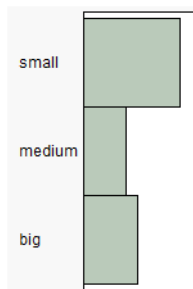
Data tables and reports are all connected in JMP. Click on a histogram bar to select the corresponding rows in the data table.

## Bar Charts

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. A bar chart shows a bar for every level of the variable, whereas the histogram shows a range of values for the variable.

---

**Figure 3.5** Example of a Bar Chart



---

### Scenario

This example uses the `Companies.jmp` data table, which contains data on the size and type of a group of companies.

A financial analyst wants to explore the following questions:

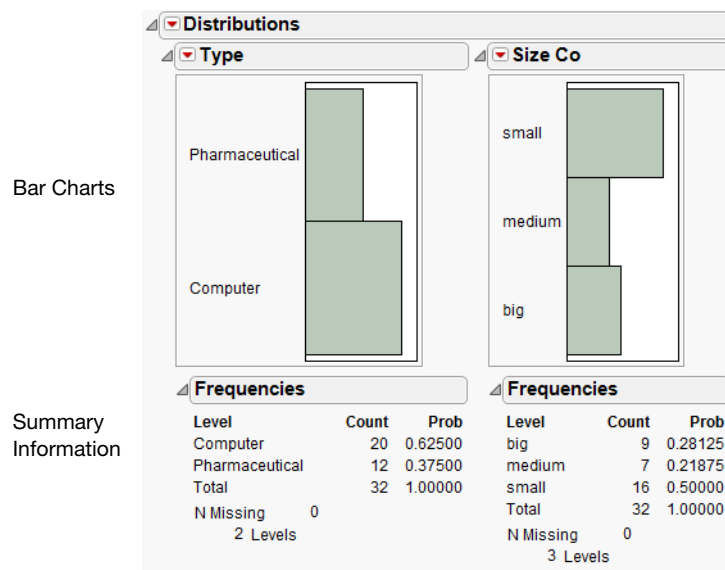
- What is the most common type of company?
- What is the most common size for a company?

To answer these questions, use bar charts of Type and Size Co.

### Creating the Bar Chart

1. Open the Companies.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Type and Size Co and click **Y, Columns**.
4. Click **OK**.

**Figure 3.6** Bar Charts of Type and Size Co



### Interpreting the Bar Charts

The bar charts provides these answers:

- There are more computer companies than pharmaceutical companies.  
The bar that represents computer companies is larger than the bar that represents pharmaceutical companies.
- The most common company size is small.  
The bar that represents small companies is larger than the bars that represent medium and big companies.

The additional summary output gives detailed frequencies. This report is discussed in “Distributions of Categorical Variables,” p. 95 in the “Analyzing Your Data” chapter.

### Interacting with the Bar Charts

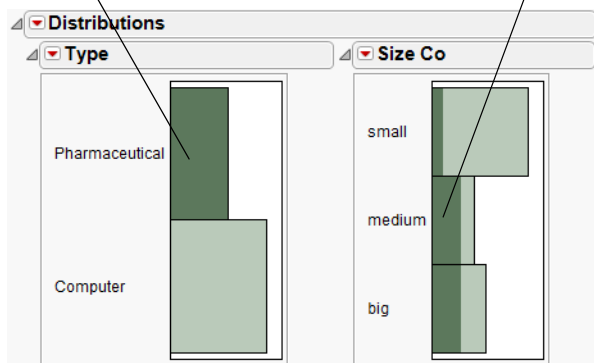
As is the case with histograms, click on individual bars to highlight rows of the data table. If more than one graph is created, clicking on a bar in one bar chart highlights the corresponding bar or bars in the other bar chart.

For example, suppose that you want to see 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. Figure 3.7 shows that although most companies in this data table are small, most of the pharmaceutical companies are medium or big.

Also, the corresponding rows in the data table are selected.

**Figure 3.7** Clicking Bars

Click on this bar to select the corresponding data in the other chart.



## Comparing Multiple Variables

Use multiple-variable graphs to visualize the relationships and patterns between two or more variables. This section covers the following graphs:

**Table 3.1** Multiple-Variable Graphs

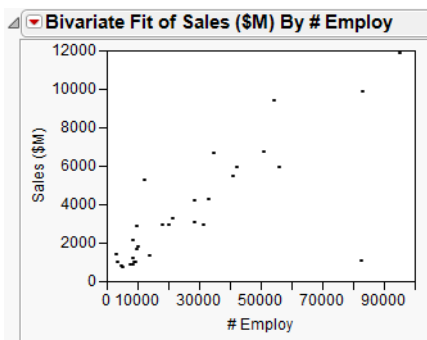
“Scatterplots,” p. 60	Use scatterplots to compare two continuous variables.
“Scatterplot Matrix,” p. 64	Use scatterplot matrices to compare several pairs of continuous variables.
“Side-by-Side Box Plots,” p. 67	Use side-by-side box plots to compare one continuous and one categorical variable.

**Table 3.1** Multiple-Variable Graphs

“Overlay Plots,” p. 69	Use overlay plots to compare one or more variables on the Y-axis to another variable on the X-axis. Overlay plots are especially useful if the X variable is a time variable, because you can compare how two or more variables change across time.
“Variability Chart,” p. 73	Use variability charts to compare one continuous Y variable to one or more categorical X variables. Variability charts show differences in means and variability across several categorical X variables.
“Graph Builder,” p. 75	Use Graph Builder to create and change graphs interactively.
“Bubble Plots,” p. 80	Bubble plots are specialized scatterplots that use color and bubble sizes to represent up to five variables at once. If one of your variables is a time variable, you can animate the plot to see your other variables change through time.

## Scatterplots

The scatterplot is the simplest of all the multiple-variable graphs. Use scatterplots to determine the relationship between two continuous variables and to discover whether two continuous variables are *correlated*. Correlation indicates how closely two variables are related. When you have two variables that are highly correlated, one might influence the other. Or, both might be influenced by other variables in a similar way.

**Figure 3.8** Example of a Scatterplot

## Scenario

This example uses the `Companies.jmp` data table, which contains sales figures and the number of employees of a group of companies.

A financial analyst wants to explore the following questions:

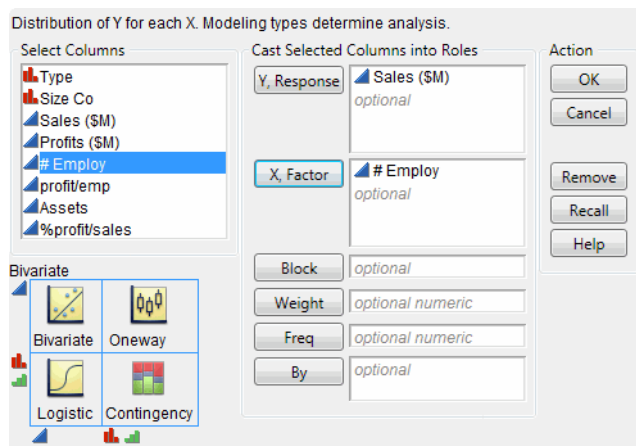
- What is the relationship between sales and the number of employees?
- Does the amount of sales increase with the number of employees?
- Can you predict average sales from the number of employees?

To answer these questions, use a scatterplot of Sales (\$M) versus # Employ.

## Creating the Scatterplot

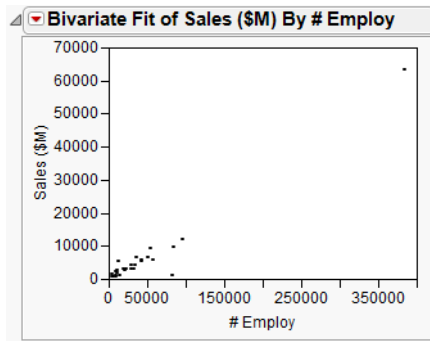
1. Open the `Companies.jmp` sample data table.
2. Select **Analyze > Fit Y by X**.
3. Select Sales (\$M) and **Y, Response**.
4. Select # Employ and **X, Factor**.

**Figure 3.9** Fit Y by X Window



5. Click **OK**.

---

**Figure 3.10** Scatterplot of Sales (\$M) versus # Employ

---

### Interpreting the Scatterplot

One company has a large number of employees and high sales, represented by the single point at the top right of the plot. The distance between this data point and all the rest makes it difficult to visualize the relationship between the rest of the companies. Remove the point from the plot and recreate the plot by following these steps:

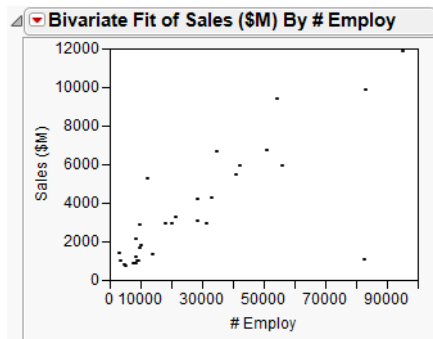
1. Click on the point to select it.
2. Select **Rows > Exclude/Unexclude**. The data point is no longer included in calculations.
3. Select **Rows > Hide/Unhide**. The data point is hidden on all graphs.

---

**Note:** The difference between hiding and excluding is important. Hiding a point removes it from any graphs but statistical calculations continue to use the point. Excluding a point removes it from any statistical calculations but does not remove it from graphs. When you both hide and exclude a point, you remove it from all calculations and from all graphs.

---

4. To recreate the plot without the outlier, select **Script > Redo Analysis** from the red triangle menu for Bivariate. You can close the original report window.

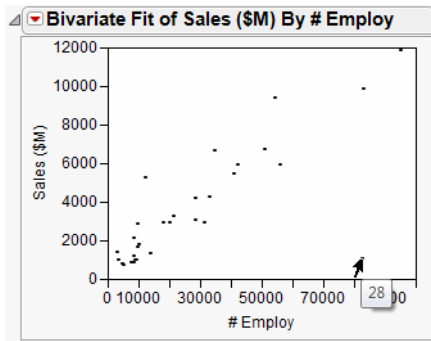
**Figure 3.11** Scatterplot with the Outlier Removed

The updated scatterplot provides these answers:

- There is a relationship between the sales and the number of employees.  
The data points have a discernible pattern. They are not scattered randomly throughout the graph. You could draw a diagonal line that would be near most of the data points.
- Sales do increase with the number of employees, and the relationship is linear.  
If you drew that diagonal line, it would slope from bottom left to top right. This slope shows that as the number of employees increases (left to right on the bottom axis), sales also increases (bottom to top on the left axis). A straight line would be near most of the data points, indicating a linear relationship. If you would have to curve your line to be near the data points, there would still be a relationship (because of the pattern of the points). However, that relationship would not be linear.
- You can predict average sales from the number of employees.  
The scatterplot shows that sales generally increase as the number of employees does. You could predict the sales for a company if you knew only the number of employees of that company. Your prediction would be on that imaginary line. It would not be exact, but it would approximate the real sales.

### ***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).

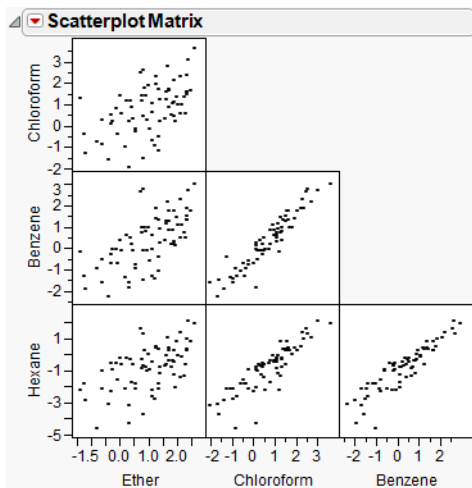
**Figure 3.12** Hover Over a Point

Click on a point to highlight the corresponding row in the data table. Select multiple points by doing one of the following:

- Click and drag with the mouse around the points. This selects points in a rectangular area.
- Select the lasso tool, and then click and drag around multiple points. The lasso tool selects an irregularly shaped area.

## 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.

**Figure 3.13** Example of a Scatterplot Matrix

### Scenario

This example uses the Solubility.jmp data table, which contains data for solubility measurements for 72 different solutes.

A lab technician wants to explore the following questions:

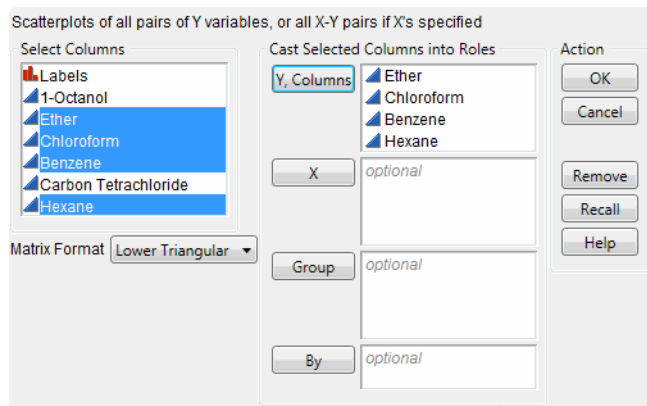
- Is there a relationship between any pair of chemicals? (There are six possible pairs.)
- Which pair has the strongest relationship?

To answer these questions, use a scatterplot matrix of the four solvents.

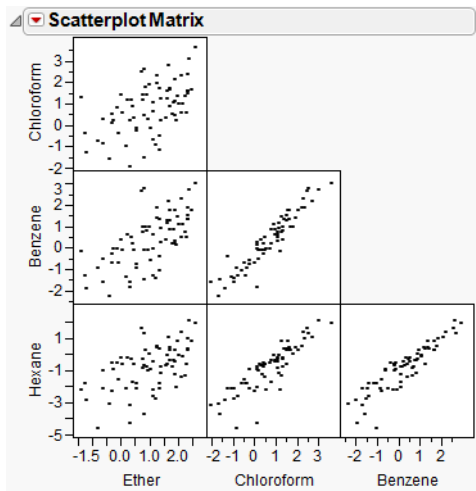
### Creating the Scatterplot Matrix

1. Open the Solubility.jmp sample data table.
2. Select **Graph > Scatterplot Matrix**.
3. Select Ether, Chloroform, Benzene, and Hexane, and click **Y, Columns**.

**Figure 3.14** Scatterplot Matrix Window



4. Click **OK**.

**Figure 3.15** Scatterplot Matrix

### ***Interpreting the Scatterplot Matrix***

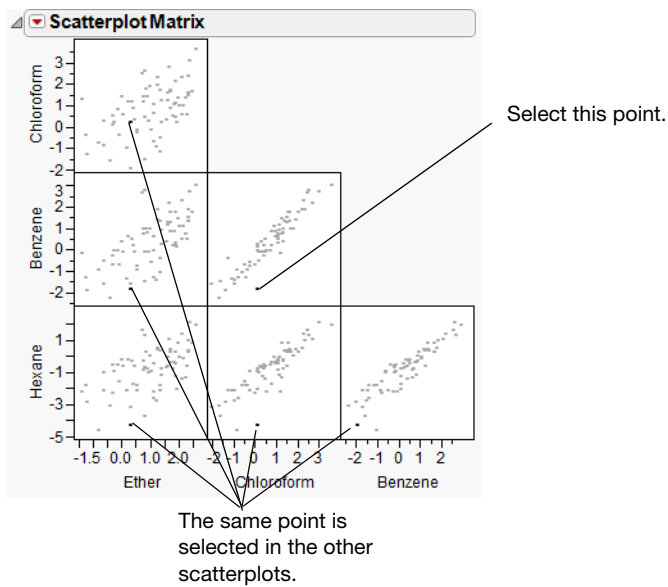
The scatterplot matrix provides these answers:

- All six pairs of variables are positively correlated.  
As one variable increases, the other variable increases too.
- The strongest relationship appears to be between Benzene and Chloroform.  
The data points in the scatterplot for Benzene and Chloroform are the most tightly clustered along an imaginary line.

### ***Interacting with the Scatterplot Matrix***

If you select a point in one scatterplot, it is selected in all the other scatterplots.

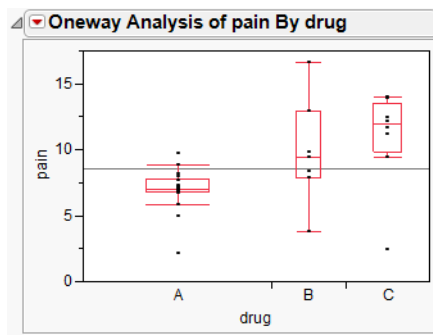
For example, if you select a point in the Benzene versus Chloroform scatterplot, the same point is selected in the other five plots.

**Figure 3.16** Selected Points

## Side-by-Side Box Plots

Side-by-side box plots show the following:

- the relationship between one continuous variable and one categorical variable
- differences in the continuous variable across levels of the categorical variable

**Figure 3.17** Example of Side-by-Side Box Plots

### Scenario

This example uses the *Analgesics.jmp* data table, which contains data on pain measurements taken on patients using three different drugs.

A researcher wants to explore the following questions:

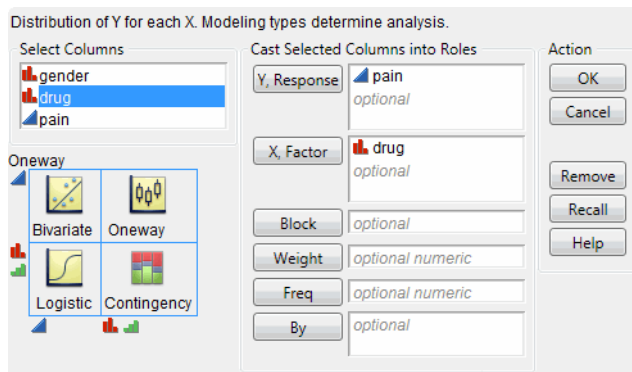
- Are there differences in the average amount of pain control among the drugs?
- Does the *variability* in the pain control given by each drug differ? A drug with high variability would not be as reliable as a drug with low variability.

To answer these questions, use a side-by-side box plot for the pain levels and the drug categories.

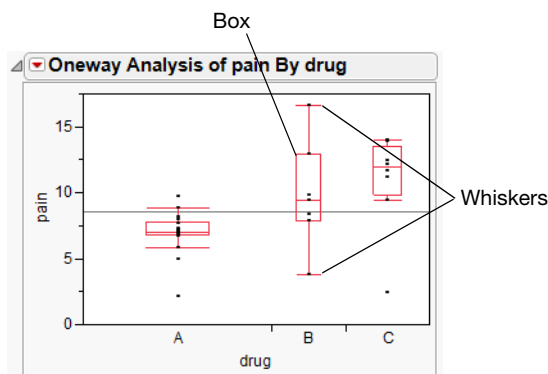
### Creating the Side-by-Side Box Plots

1. Open the *Analgesics.jmp* data table.
2. Select **Analyze > Fit Y by X**.
3. Select *pain* and click **Y, Response**.
4. Select *drug* and click **X, Factor**.

**Figure 3.18** Fit Y by X Window



5. Click **OK**.
6. From the red triangle menu, select **Display Options > Box Plots**.

**Figure 3.19** Side-by-Side Box Plots

### Interpreting the Side-by-Side Box Plots

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 box.
- The majority of the data falls between the ends of the whiskers.
- A data point outside the whiskers might be an outlier.

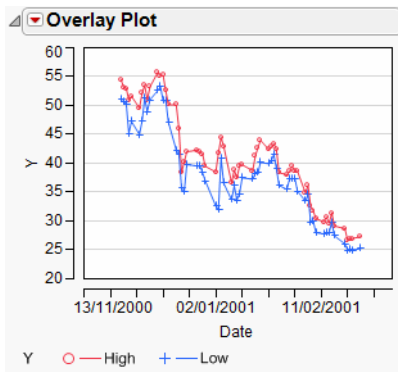
The box plots in Figure 3.19 show these answers:

- There is evidence to believe that patients on drug A feel less pain, since the box plot for drug A is lower on the pain scale 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 points for drug C. Hover over it with your mouse 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.

### Overlay Plots

Like scatterplots, overlay plots show the relationship between two or more variables. However, if one of the variables is a time variable, an overlay plot shows trends across time better than scatterplots do.

**Figure 3.20** Example of an Overlay Plot

**Note:** To plot data over time, you can also use Graph Builder, bubble plots, control charts, and variability charts. For complete details, see *Basic Analysis and Graphing* or *Quality and Reliability Methods*.

## Scenario

This example uses the `Stock Prices.jmp` data table, which contains data on the price of a stock over a three month period.

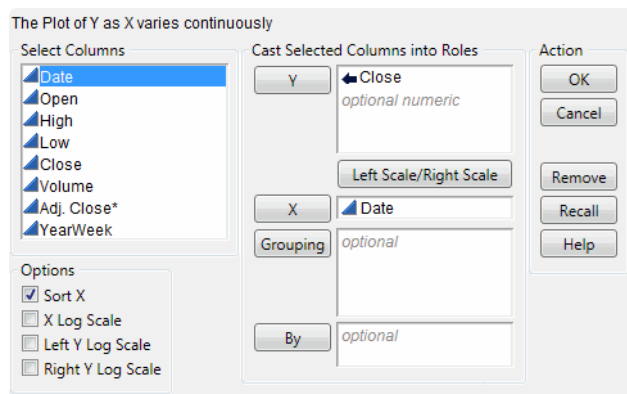
A potential investor wants to explore the following questions:

- Has the stock's closing price changed over the last three months?  
To answer this question, use an overlay plot of the stock's closing price over time.
- How do the stock's high and low prices relate to each other?  
To answer this question, use another overlay plot of the stock's high and low prices over time.

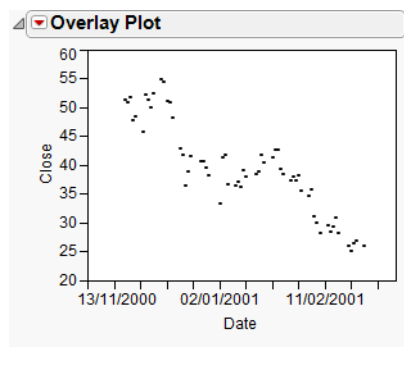
Create the first overlay plot to answer the first question, and then create a second overlay plot to answer the second question.

## Creating the Overlay Plot of the Stock's Price over Time

1. Open the `Stock Prices.jmp` data table.
2. Select **Graph > Overlay Plot**.
3. Select **Close** and click **Y**.
4. Select **Date** and click **X**.

**Figure 3.21** Overlay Plot Window

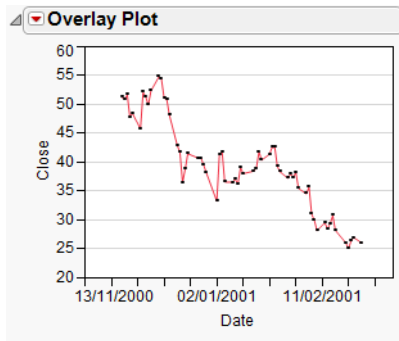
5. Click **OK**.

**Figure 3.22** Overlay Plot of the Closing Price Over Time

### ***Interpreting and Interacting with the Overlay Plot***

The overlay plot shows that the closing stock price has been decreasing over the last several months. To see the trend more clearly, connect the points and add grid lines.

1. From the red triangle menu, select **Connect Thru Missing**.
2. Double-click on the Y axis.
3. Select the **Major Gridline** check box.
4. Click **OK**.

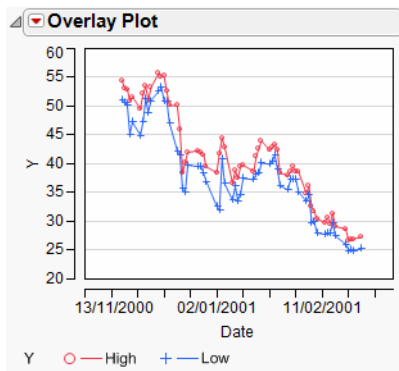
**Figure 3.23** Connected Points and Grid Lines


The potential investor can see that although the stock price has gone up and down over the last three months, the overall trend has been downward.

### Creating the Overlay Plot of the Stock's High and Low Prices

Use an overlay plot to plot more than one Y variable. For example, suppose that you want to see both the high and the low prices on the same plot.

1. Follow the steps in “Creating the Overlay Plot of the Stock's Price over Time,” p. 70, this time assigning both High and Low to the Y role.
2. Connect the points and add grid lines as shown in “Interpreting and Interacting with the Overlay Plot,” p. 71.

**Figure 3.24** Two Y Variables


The legend at the bottom of the plot shows the colors and markers used for the High and Low variables in the graph. The overlay plot shows that the High price and Low price track each other very closely.

## Answering the Questions

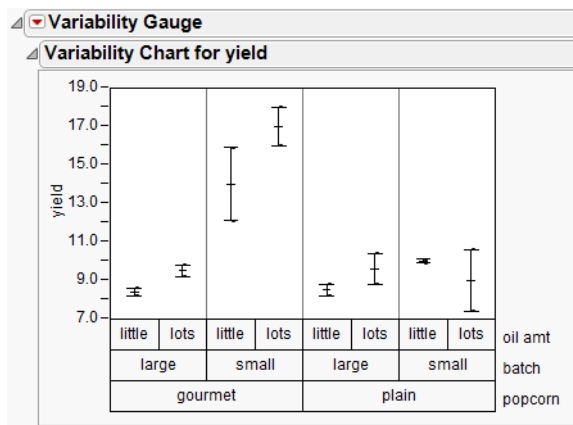
Both of the overlay plots answer the two questions asked at the beginning of this example.

- The first plot shows that the price of this stock has not remained the same, but has been decreasing.
- The second plot shows that the high and low price of this stock are not very different from each other. The stock price does not vary wildly on any given day.

## Variability Chart

In the graphs described so far, you specified only a single X variable. Use a variability chart to specify multiple X variables and see differences in means and variability across all of your variables at once.

**Figure 3.25** Example of a Variability Chart



## Scenario

This example uses the Popcorn.jmp data table with data from a popcorn maker. The yield (the volume of popcorn for a given measure of kernels) was measured for each combination of popcorn style, batch size, and amount of oil used.

The popcorn maker wants to explore the following question:

- Which combination of factors results in the highest popcorn yield?

To answer this question, use a variability chart of the yield versus the style, batch size, and oil amount.

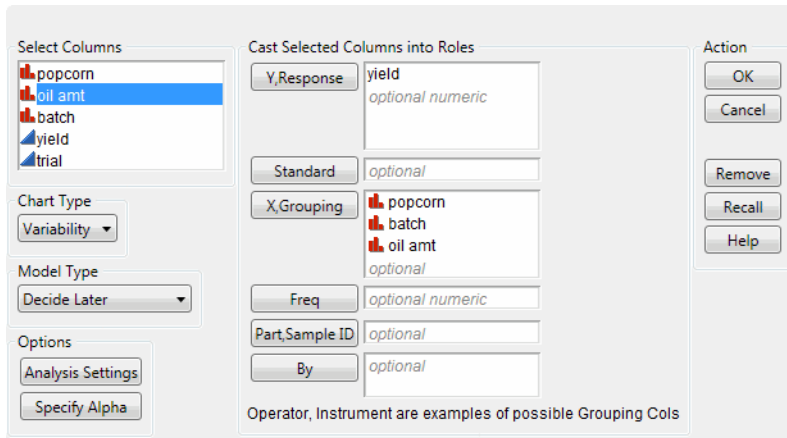
## Creating the Variability Chart

1. Open the Popcorn.jmp data table.
2. Select **Graph > Variability/Gauge Chart**.
3. Select yield and click **Y, Response**.

4. Select popcorn and click **X, Grouping**.
5. Select batch and click **X, Grouping**.
6. Select oil amt and click **X, Grouping**.

**Note:** The order in which you assign the variables to the **X, Grouping** role is important, because the order in this window determines their nesting order in the variability chart.

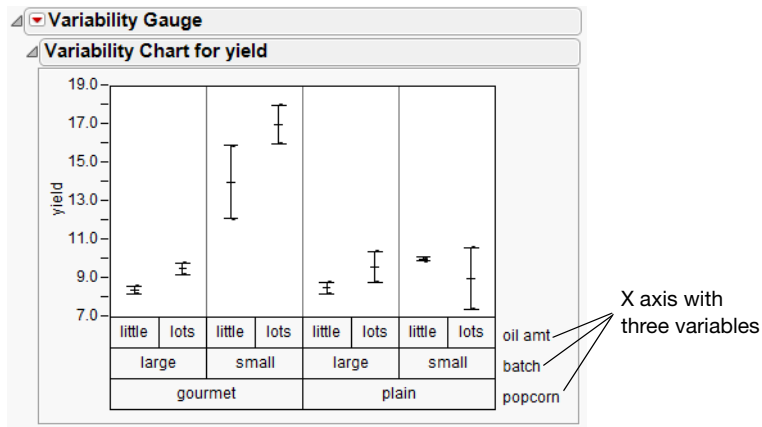
**Figure 3.26** Variability Chart Window



7. Click **OK**.

The top chart is the variability chart, showing 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. Since the bottom chart does not show the yield, hide it.

8. Deselect **Std Dev Chart** on the red triangle menu.

**Figure 3.27** Results Window

### Interpreting the Variability Chart

The variability chart for yield indicates that small, gourmet batches produce the highest yield.

To be more specific, the popcorn maker might ask this additional question: Is the yield high because those batches are small, or because those batches are gourmet?

The variability chart shows the following:

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

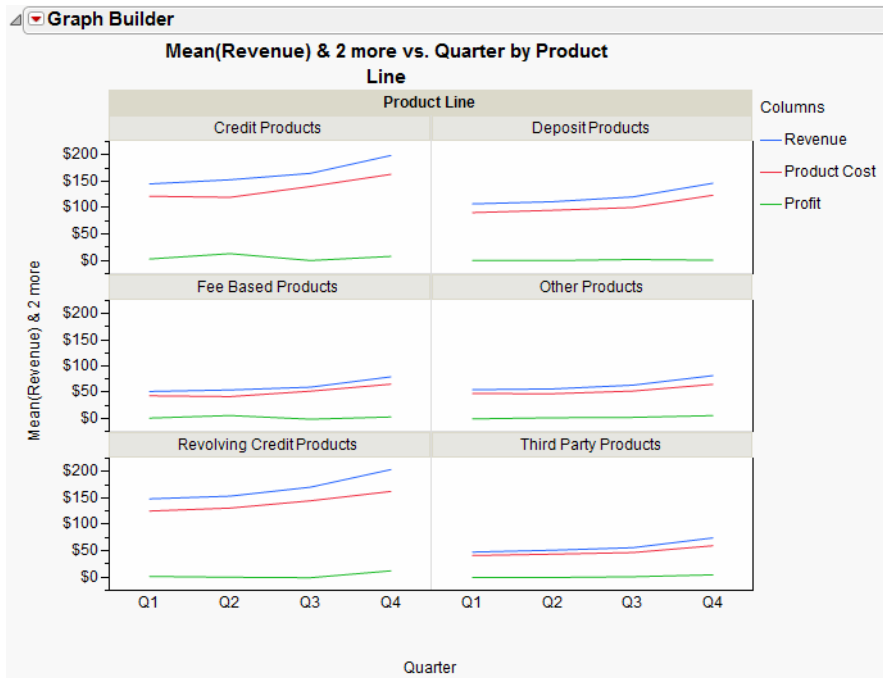
Given this information, the popcorn maker can conclude that only the combination of small and gourmet at the same time results in batches with high yield. It would have been impossible to reach this conclusion with a chart that only allowed a single variable.

### Graph Builder

Use Graph Builder to interactively create and modify graphs. So far, all of the graphs have been created by launching a platform and specifying variables. Once the graph is created, you cannot change the variables. To create a different type of graph, you must launch a different platform. In Graph Builder, you can change the variables and change the graphs at any time.

Use Graph Builder to accomplish the following tasks:

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

**Figure 3.28** Example of a Graph that was Created with Graph Builder


**Note:** Only some of the Graph Builder features are covered here. For complete details, see *Basic Analysis and Graphing*.

### Scenario

This example uses the Profit by Product.jmp data table, which contains profit data for multiple product lines.

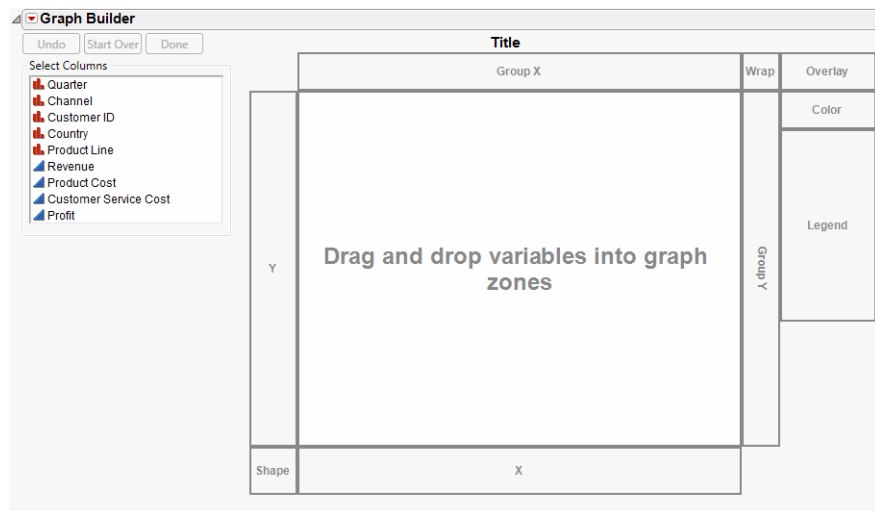
A business analyst wants to explore the following question:

- How is the profitability different between product lines?

To answer this question, use a line plot that displays revenue, product cost, and profit data across different product lines.

### Creating the Graph

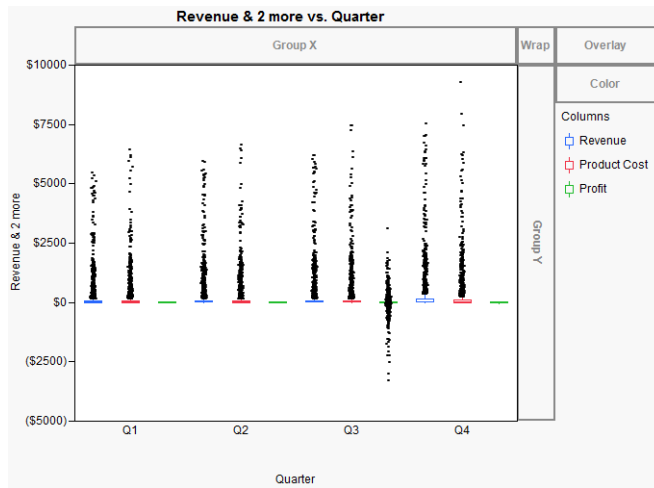
1. Open the Profit by Product.jmp data table.
2. Select **Graph > Graph Builder**.

**Figure 3.29** Graph Builder Workspace

3. Click on **Quarter** and then drag and drop it onto the **X** zone to assign **Quarter** as the **X** variable.
4. Click on **Revenue**, **Product Cost**, and **Profit**, and drag and drop them onto the **Y** zone to assign all three variables as **Y** variables.

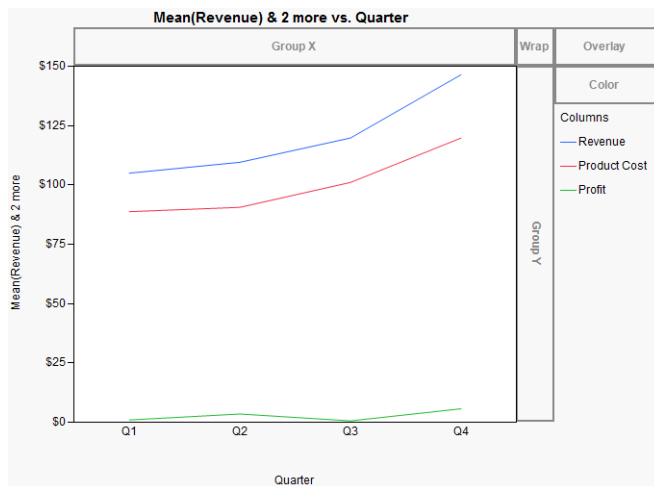
The **X** and **Y** zones are now axes.

**Note:** You can also click on variables and then click a zone to assign them. However, after a zone becomes an axis, drag and drop additional variables onto the axis rather than clicking on the variables and axis.

**Figure 3.30** After Adding Y and X Variables


Based on the variables that you are using, Graph Builder shows side-by-side box plots.

- To change the box plots to a line plot, right-click on the plot and select **Box Plot > Change To > Line**.

**Figure 3.31** Line Plot


- To create a separate chart for each product, click on **Product Line**, and drag and drop it into the **Wrap** zone.

A separate line plot is created for each product.

**Figure 3.32** Final Line Plots



### Interpreting the Graph

Figure 3.32 shows revenue, cost, and profit broken down by product line. The business analyst was interested in seeing the difference in profitability between product lines. The line plots in Figure 3.32 can provide some answers, as follows:

- Credit products, deposit products, and revolving credit products produce more revenue than fee-based products, third-party products, and other products.
- However, the profits of all the product lines are similar.

The data table also includes data on sales channels. The business analyst wants to see how revenue, product cost, and profit differ between different sales channels.

- To remove **Product Line** from the graph, click the title of the graph (**Product Line**) and drag and drop it into any empty area within Graph Builder.
- To add **Channel** as the wrap variable, click on **Channel** and drag and drop it into the **Wrap** zone.

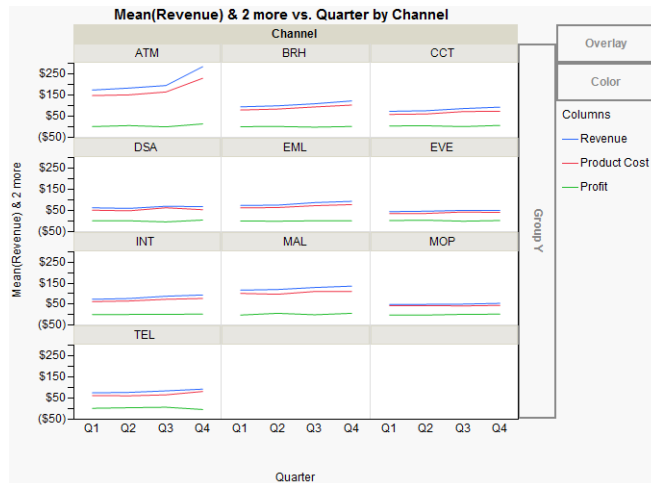
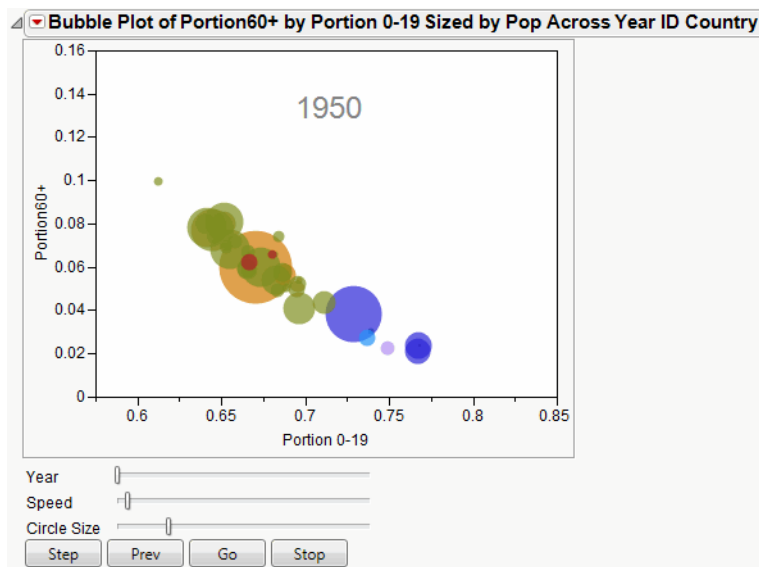
**Figure 3.33** Line Plots Showing Sales Channels


Figure 3.33 provides this answer: revenue and product cost for ATMs are the highest and are growing the most quickly.

## Bubble Plots

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.

**Figure 3.34** Example of a Bubble Plot

### Scenario

This example uses the `PopAgeGroup.jmp` data table, which contains population statistics 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.

A sociologist wants to explore the following question:

- Is the age of the population of the world changing?

To answer this question, look at the relationship between the oldest (over 59) and the youngest (younger than 20) portions of the population. Use a bubble plot to determine how this relationship changes over time.

### Creating the Bubble Plot

1. Open the `PopAgeGroup.jmp` data table.
2. Select **Graph > Bubble Plot**.
3. Select `Portion60+` and click **Y**.  
This corresponds to the *Y* variable on the bubble plot.
4. Select `Portion 0-19` and click **X**.  
This corresponds to the *X* variable on the bubble plot.

5. Select `Country` and click **ID**.

Each unique level of the ID variable is represented by a bubble on the plot.

6. Select Year and click **Time**.

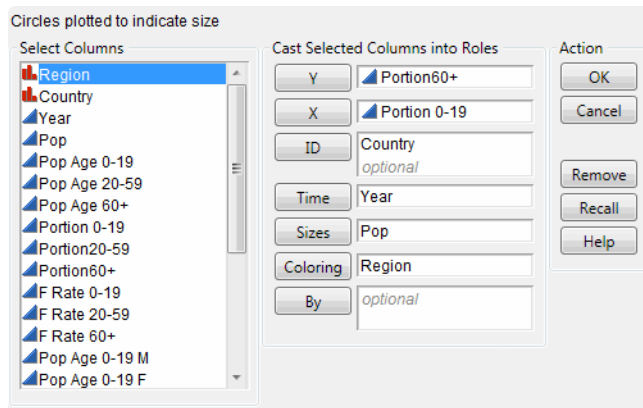
This controls the time indexing when the bubble plot is animated.

7. Select Pop and click **Sizes**.

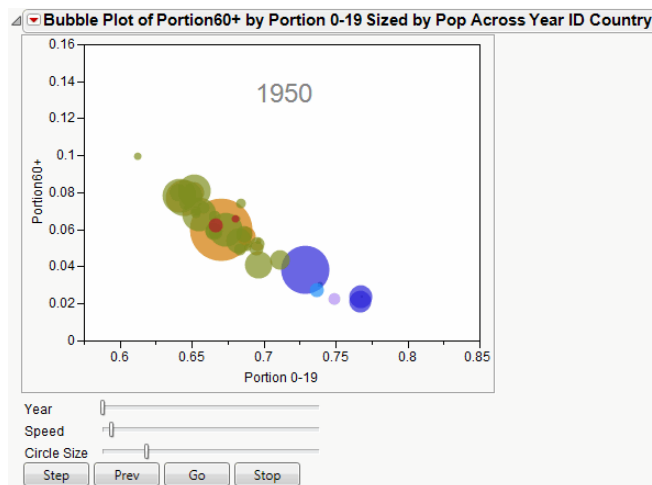
This controls the size of the bubbles.

8. Select Region and click **Coloring**.

Each unique level of the Coloring variable is reflected by a bubble of that color. So in this example, all the bubbles for countries located in the same region are the same color.

**Figure 3.35** Bubble Plot Window


9. Click **OK**.

**Figure 3.36** 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. Animate the bubble plot to cycle through all the years by clicking **Go**. 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 *Basic Analysis and Graphing*.

---

The bubble plot for 1950 shows that if a country's proportion of people younger than 20 is high, then the proportion of people over 59 is low.

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

**Stop** stops the animation.

**Step** manually controls 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.

The sociologist wanted to know how the age of the world's population is changing. The bubble plot indicates that the population of the world is getting older.

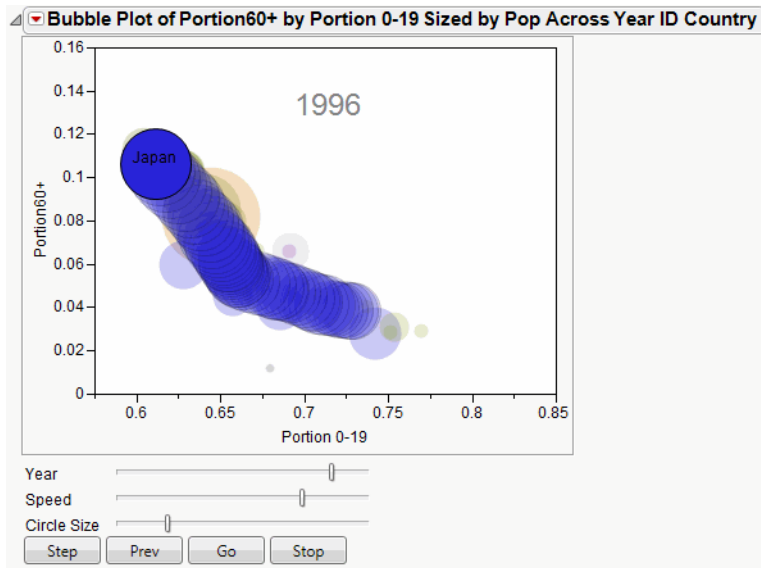
### Interacting with the Bubble Plot

Click to select a bubble to see the trend for that bubble over time. For example, in the 1950 plot, the large bubble in the middle is Japan.

To see the pattern of population changes in Japan through the years

1. Click in the middle of the Japan bubble to select it.
2. From the red triangle menu, select **Trail Bubbles**.
3. Click **Go**.

As the animation progresses through time, the Japan bubble leaves a trail of bubbles that illustrates its history.

**Figure 3.37** Japan's History of Population Shifts

Focusing on the Japan bubble, you can see the following over time:

- The proportion of the population 19 years old or less decreased.
- The proportion of the population 60 years old or more increased.

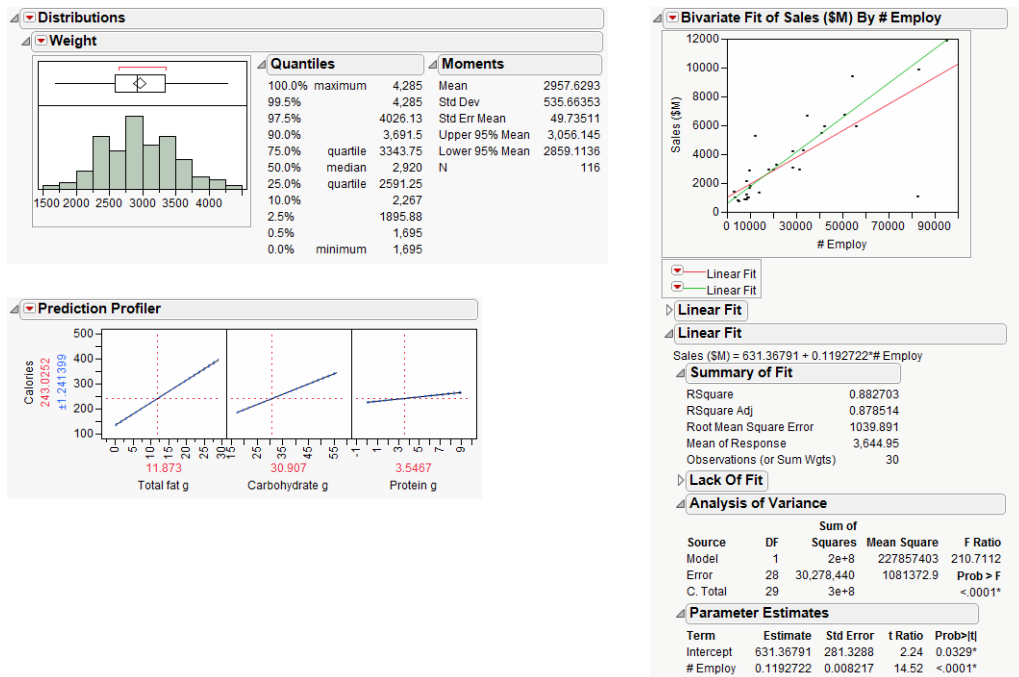
# Chapter 4

## Analyzing Your Data Distributions, Relationships, and Models

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

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

Figure 4.1 Analysis Examples



# Contents

About This Chapter . . . . .	87
The Importance of Graphing Your Data . . . . .	87
Understanding Modeling Types. . . . .	89
Example: Modeling Type Results. . . . .	90
Changing the Modeling Type . . . . .	91
Analyzing Distributions. . . . .	93
Distributions of Continuous Variables. . . . .	93
Distributions of Categorical Variables . . . . .	95
Analyzing Relationships. . . . .	98
Using Regression with One Predictor . . . . .	99
Comparing Averages for One Variable. . . . .	103
Comparing Proportions. . . . .	105
Comparing Averages for Multiple Variables . . . . .	108
Using Regression with Multiple Predictors . . . . .	112

---

## About This Chapter

Before you analyze your data, review the following information:

- “The Importance of Graphing Your Data,” p. 87
- “Understanding Modeling Types,” p. 89

The rest of this chapter shows you how to use some basic analytical methods in JMP:

- “Analyzing Distributions,” p. 93
- “Analyzing Relationships,” p. 98

For a description of advanced modeling and analysis techniques, see *Modeling and Multivariate Methods* and *Quality and Reliability Methods*.

---

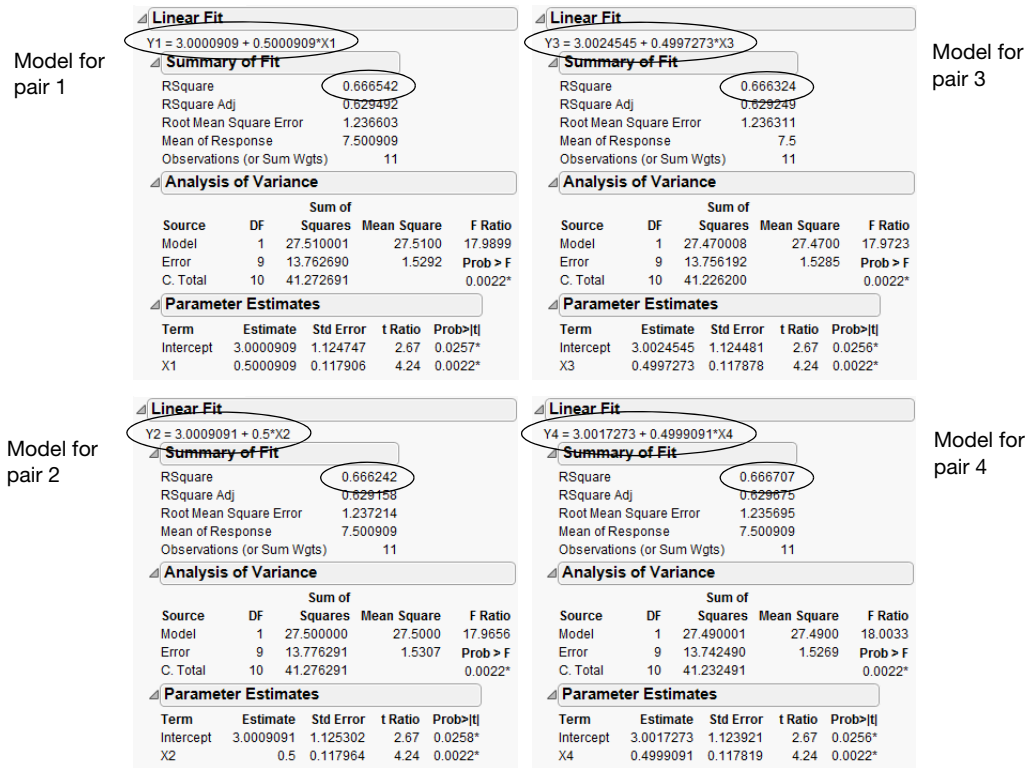
## The Importance of Graphing Your Data

Graphing, or visualizing, your data 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 (E. J. Anscombe (1973), *American Statistician*, 27, 17-21). This data consists of four pairs of  $X$  and  $Y$  variables.
2. From the red triangle menu for the **The Quartet** script in the table panel, 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.

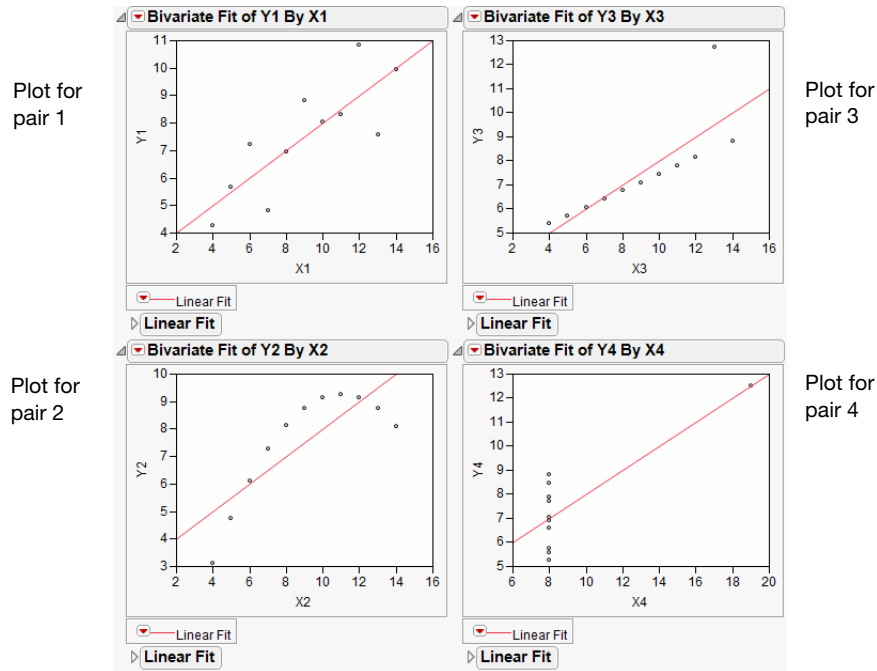
Figure 4.2 Four Models



Notice that all four models and the 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. From the red triangle menu for any one of the Bivariate Fits, select **Show Points**.

**Figure 4.3** Scatterplots with Points Added

The scatterplots show that the relationship between  $X$  and  $Y$  is not the same for the four pairs, although the lines describing the relationships are the same:

- Plot 1 represents a 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

In JMP, data can be of different types. JMP refers to this as the modeling type of the data. Table 4.1 describes the three modeling types in JMP.

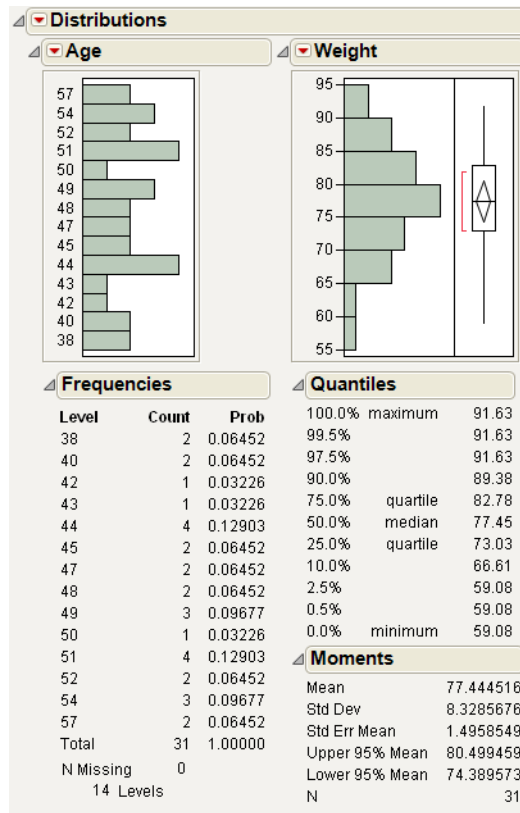
**Table 4.1** Modeling Types

Modeling Type	Description	Examples	Specific Example
Continuous	Numeric data only. Used in operations like sums and means.	Height Temperature Time	The time to complete a test might be 2 hours, or 2.13 hours.
Ordinal	Numeric or character data. Values belong to ordered categories.	Month (1,2,...,12) Letter grade (A, B,...F) Size (small, medium, large)	The month of the year can be 2 (February) or 3 (March), but not 2.13. February comes before March.
Nominal	Numeric or character data. Values belong to categories, but the order is not important.	Gender (M or F) Color Test result (pass or fail)	The gender can be M or F, with no order. Gender categories can also be represented by a number (M=1 and F=2).

### Example: Modeling Type Results

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

1. Open the Linnerud.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Age and Weight and click **Y, Columns**.
4. Click **OK**.

**Figure 4.4** Distribution Results for Age and Weight

Although Age and Weight are both numeric variables, they are not treated the same. Table 4.2 compares the differences between the results for weight and age.

**Table 4.2** Results for weight and age

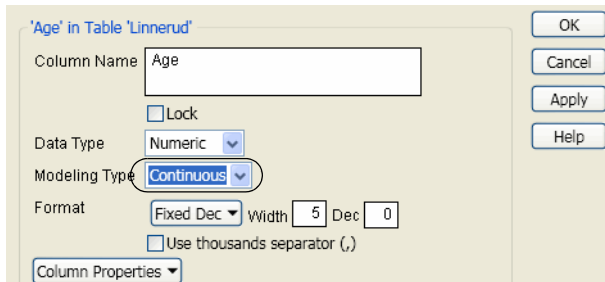
Variable	Modeling Type	Results
Weight	Continuous	Histogram, Quantiles, and Moments
Age	Ordinal	Bar chart and Frequencies

## Changing the Modeling Type

To treat a variable differently, change the modeling type. For example, in Figure 4.4, the modeling type for Age is ordinal. Remember that for an ordinal variable, JMP calculates frequency counts. Suppose that you wanted to find the average age instead of frequency counts. Change the modeling type to continuous, which shows the mean age.

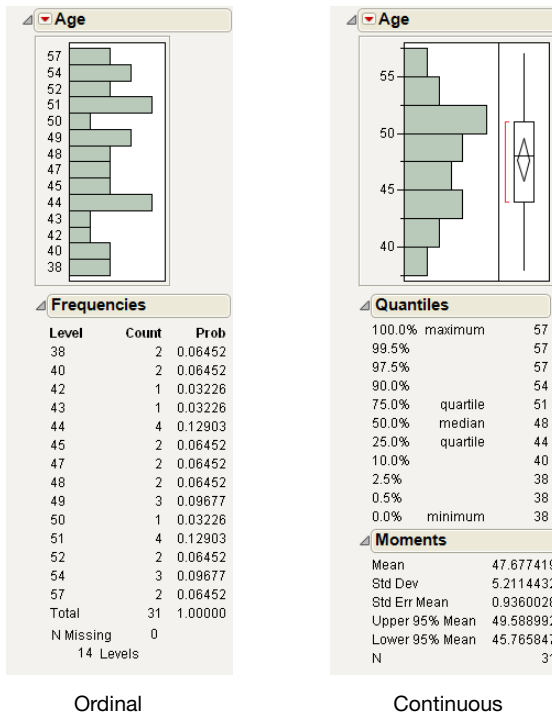
1. Double-click on the Age column heading. The Column Info window appears.
2. Change the Modeling Type to **Continuous**.

**Figure 4.5** Column Info Window



3. Click **OK**.
4. Repeat the steps in the example (see “Example: Modeling Type Results,” p. 90) to create the distribution. Figure 4.6 shows the distribution results when Age is ordinal and continuous.

**Figure 4.6** Different Modeling Types for age



When age is ordinal, you can see the frequency counts for each age. For example, age 48 appears 2 times. When age is continuous, you can find the mean age, which is nearly 48 (47.677)

---

## Analyzing Distributions

To analyze a single variable, you can examine the distribution of the variable, using the Distribution platform. Report content for each variable varies, depending on whether the variable is categorical (nominal or ordinal) or continuous.

---

**Note:** For complete details about the **Distribution** platform, see *Basic Analysis and Graphing*.

---

### Distributions of Continuous Variables

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

- Does the shape of the data match any known distributions?
- Are there any outliers in the data?
- 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?

You can answer these and other questions with graphs, summary statistics, and simple statistical tests.

#### Scenario

This example uses the *Car Physical Data.jmp* data table, which contains information about 116 different car models.

A planning specialist has been asked by a railroad company to determine the possible issues involved in transporting cars by train. Using the data, the planning specialist wants to explore the following questions:

- What is the average car weight?
- How spread out are the cars' weights (standard deviation)?
- What are the minimum and maximum weights of cars?
- Are there any outliers in the data?

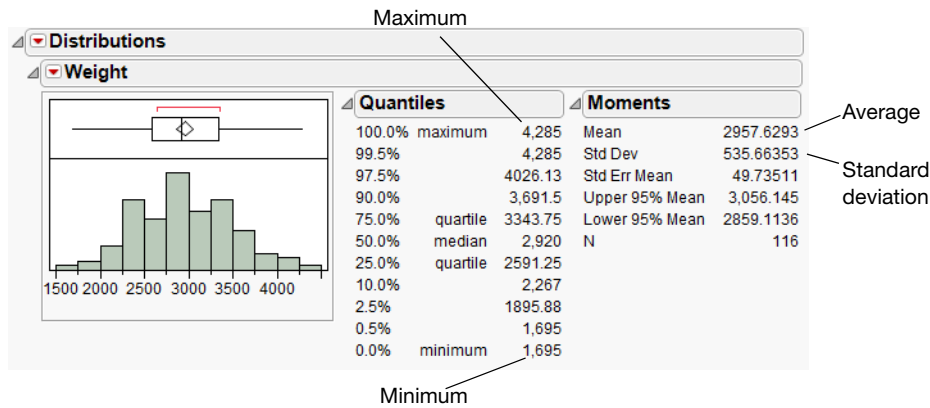
Use a histogram of weight to answer these questions.

#### Creating the Histogram

1. Open the *Car Physical Data.jmp* sample data table.
2. Select **Analyze > Distribution**.

3. Select **Weight** and click **Y, Columns**.
4. Click **OK**.
5. To rotate the report window, select **Display Options > Horizontal Layout** from the red triangle menu next to **Weight**.

**Figure 4.7** Distribution of Weight



The report window contains three sections:

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

**Interpreting the Distribution Results**

Using the results presented in Figure 4.7, the planning specialist can answer the questions.

**Table 4.3** Questions Answered Using the Report Window Results<sup>a</sup>

Question	Histogram	Quantiles	Moments
What is the average car weight?	about 3000 lbs		2958 lbs
How spread out are the weights (standard deviation)?			536 lbs
What are the minimum and maximum weights?	about 1500 lbs and 4500 lbs	1695 lbs and 4285 lbs	
Are there any outliers?	No		

a. A blank cell means not applicable.

The default report window in Figure 4.7 provides a minimal set of graphs and statistics. Additional graphs and statistics are available on the red triangle menu.

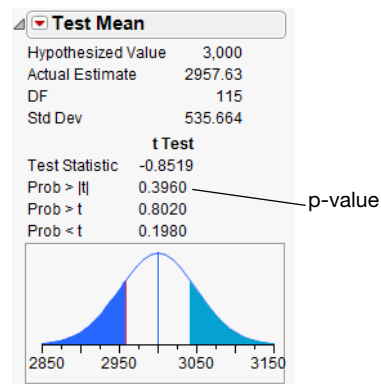
### Drawing Conclusions

Based on other research, the railroad company has determined that an average weight of 3000 pounds is the most efficient to transport. Now, the planning specialist needs to find out whether the average car weight in the general population of cars that they might transport is 3000 pounds. Use a t-test to draw inferences about the broader population based on this sample of the population.

### Testing Conclusions

1. From the red triangle menu for **Weight**, select **Test Mean**.
2. In the window that appears, type 3000 in the Specify Hypothesized Mean box.
3. Click **OK**.

**Figure 4.8** Test Mean Results



### Interpreting the t-Test

The primary result of a t-test is the p-value. In this example, the p-value is 0.396 and the analyst is using a significance level of 0.05. Since 0.396 is greater than 0.05, you cannot conclude that the average weight of car models in the broader population is significantly different from 3000 pounds. Had the p-value been lower than the significance level, the planning specialist would have concluded that the average car weight in the broader population *is* significantly different from 3000 pounds.

## Distributions of Categorical Variables

Analyzing a categorical (ordinal or nominal) variable might include questions such as the following:

- How many levels does the variable have?

- How many data points does each level have?
- Is the data uniformly distributed?
- What proportions of the total do each level represent?

### Scenario

See the scenario in “Distributions of Continuous Variables,” p. 93.

Now that the railroad company has determined that the average weight of the cars is not significantly different from the target weight, the railroad company has more questions about the cars.

The planning specialist wants to answer these questions for the railroad company:

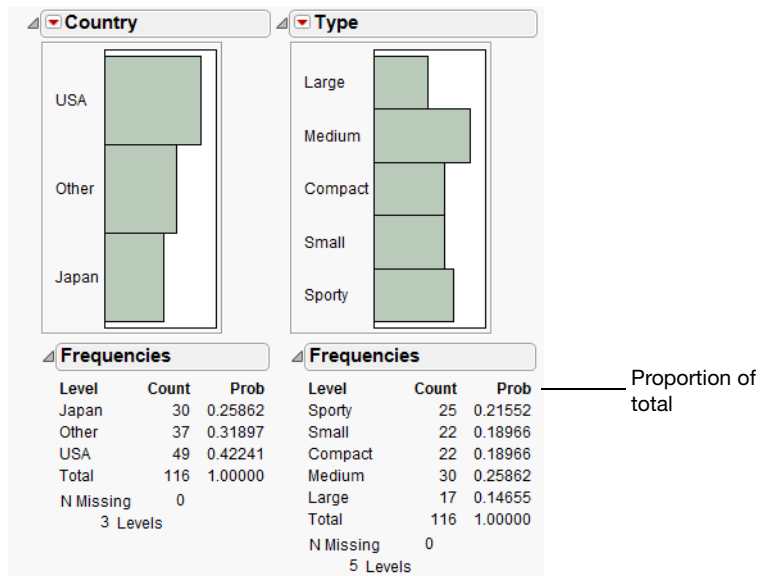
- What are the types of cars?
- What are the countries of origin?

To answer these questions, look at the distribution for Type and Country.

### Creating the Distribution

1. Open the Car Physical Data.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Country and Type and click **Y, Columns**.
4. Click **OK**.

**Figure 4.9** Distribution for Country and Type



### Interpreting the Distribution Results

The report window includes a bar chart and a Frequencies report for country and type. The bar chart is a graphical representation of the frequency information provided in the Frequencies report. The Frequencies report contains the following:

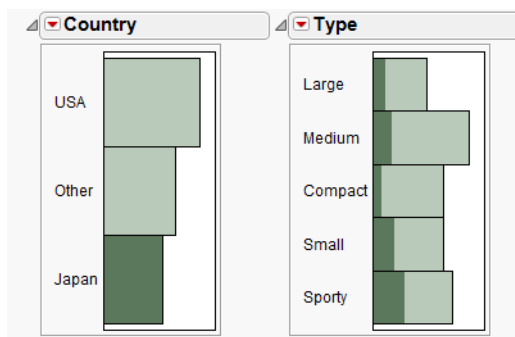
- Categories of data. For example, Japan is a category of Country, and Sporty is a category of Type.
- Total counts for each category.
- Proportion of the total each category represents.

For example, there are 22 compact cars, or about 19% of the 116 observations.

### Interacting with the Distribution Results

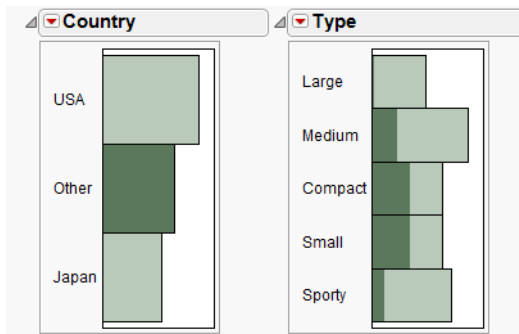
Selecting a bar in one chart also selects the corresponding data in the other chart. For example, select the Japan bar in the Country bar chart to see that a large number of Japanese cars are sporty.

**Figure 4.10** Japanese Cars



Select the Other category to see that a majority of these cars are small or compact, and almost none are large.

Figure 4.11 Other Cars



## Analyzing Relationships

Scatterplots and other such graphs can help you visualize relationships between variables. Once you have visualized relationships, the next step is to analyze those relationships so that you can describe them numerically. That numerical description of the relationship between variables is called a *model*. Even more importantly, a model also predicts the average value of one variable ( $Y$ ) from the value of another variable ( $X$ ). The  $X$  variable is also called a predictor. Generally, this model is called a *regression* model.

With JMP, the **Fit Y by X** platform and the **Fit Model** platform creates regression models.

**Note:** Only the basic platforms and options are covered here. For complete details and explanations of all platform options, see *Basic Analysis and Graphing*, *Modeling and Multivariate Methods*, and *Quality and Reliability Methods*.

Table 4.4 shows the four primary types of relationships.

**Table 4.4** Relationship Types

X	Y	Section
Continuous	Continuous	<ul style="list-style-type: none"> <li>“Using Regression with One Predictor,” p. 99</li> <li>“Using Regression with Multiple Predictors,” p. 112</li> </ul>
Categorical	Continuous	<ul style="list-style-type: none"> <li>“Comparing Averages for One Variable,” p. 103</li> <li>“Comparing Averages for Multiple Variables,” p. 108</li> </ul>
Categorical	Categorical	“Comparing Proportions,” p. 105
Continuous	Categorical	Logistic regression is an advanced topic. See <i>Basic Analysis and Graphing</i> .

## Using Regression with One Predictor

### Scenario

This example uses the `Companies.jmp` data table, which contains financial data for 32 companies from the pharmaceutical and computer industries.

Intuitively, it makes sense that companies with more employees can generate more sales revenue than companies with fewer employees. A data analyst wants to predict the overall sales revenue for each company based on the number of employees.

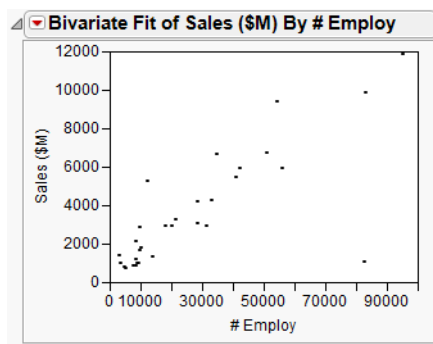
To accomplish this task, do the following:

- “Discovering the Relationship,” p. 99
- “Fitting the Regression Model,” p. 100
- “Predicting Average Sales,” p. 101

### Discovering the Relationship

First, create a scatterplot to see the relationship between the number of employees and the amount of sales revenue. This scatterplot was created in “Creating the Scatterplot,” p. 61 in the “Visualizing Your Data” chapter. After hiding and excluding one outlier (a company with significantly more employees and higher sales), the plot in Figure 4.12 shows the result.

**Figure 4.12** Scatterplot of Sales (\$M) versus # Employ

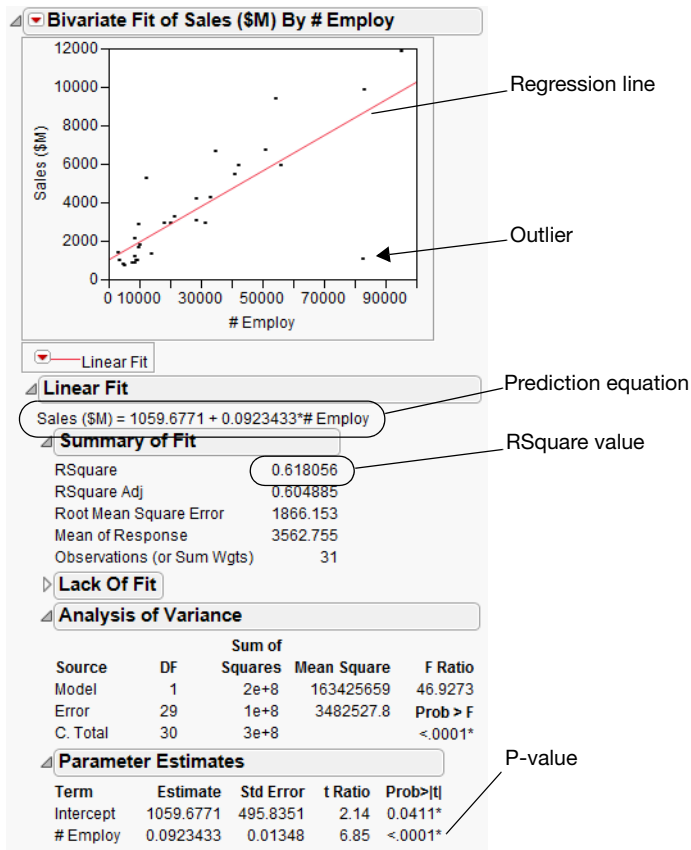


This scatterplot provides a clearer picture of the relationship between sales and the number of employees. As expected, the more employees a company has, the higher sales it can generate. This visually confirms the data analyst’s guess, but it does not predict sales for a given number of employees.

### Fitting the Regression Model

To predict the sales revenue from the number of employees, fit a regression model. From the red triangle for **Bivariate Fit**, select **Fit Line**. A regression line is added to the scatterplot and reports are added to the report window.

**Figure 4.13** Regression Line



Within the reports, look at the following results:

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

From these results, the data analyst can conclude the following:

- The p-value is less than the significance level of 0.05. Therefore, including the number of employees in the prediction model significantly improves the ability to predict average sales.

- Since the RSquare value in this example is large, this confirms that a prediction model based on the number of employees can predict sales revenue. The RSquare value shows the strength of a relationship between variables, also called the *correlation*. A correlation of 0 indicates no relationship between the variables, and a correlation of 1 indicates a perfect linear relationship.

### **Predicting Average Sales**

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 included in the report:

$$\text{Average sales} = 1059.68 + 0.092 * \text{employees}$$

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

$$\$7,499.68 = 1059.68 + 0.092 * 70,000$$

In the lower right area of the current scatterplot, there is an outlier that does not follow the general pattern of the other companies. The data analyst wants to know whether the prediction model will change when this outlier is excluded.

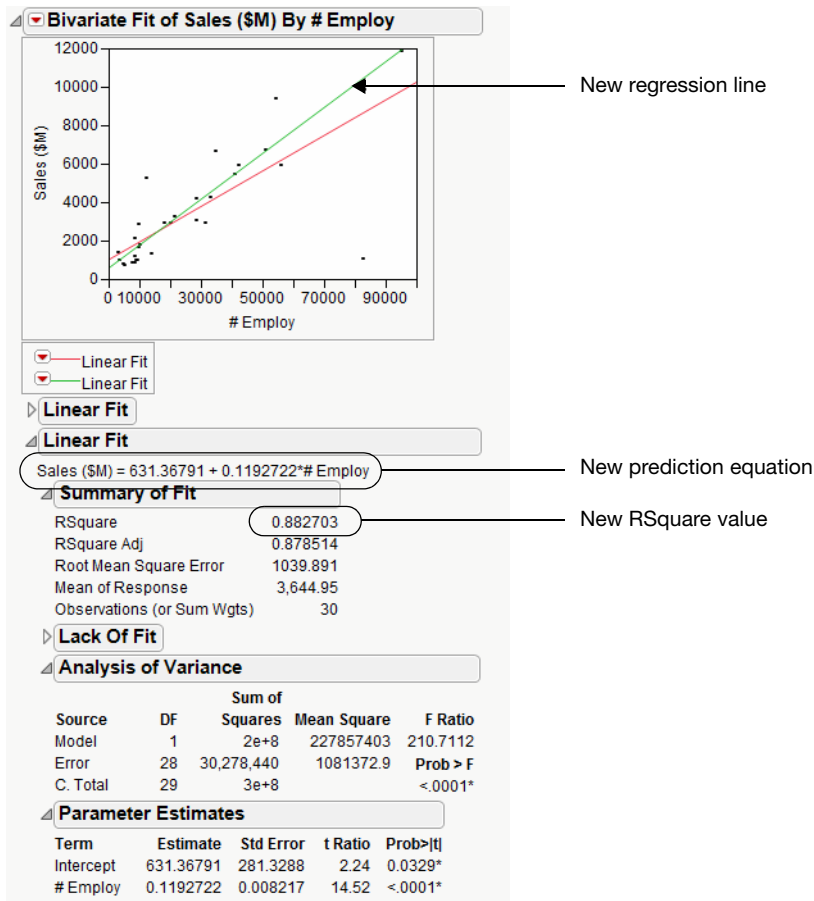
#### **Exclude the outlier**

1. Click on the outlier.
2. Select **Rows > Exclude/Unexclude**.
3. Fit this model by selecting **Fit Line** from the red triangle menu for **Bivariate Fit**.

The following are added to the report window (see Figure 4.14):

- a new regression line
- a new Linear Fit report, which includes:
  - a new prediction equation
  - a new RSquare value

Figure 4.14 Comparing the Models



Using the results in Figure 4.14, the data analyst 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 RSquare value (0.88) is closer to 1 than the first RSquare value (0.618).

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

$$\$8961.37 = 631.37 + 0.119 \times 70,000$$

The prediction for the first model was \$7499.68, so this model predicts a higher sales total by \$1461.69.

The second model, after removing the outlier, describes and predicts sales totals based on the number of employees better than the first model. The data analyst now has a good model to use.

## Comparing Averages for One Variable

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

### Scenario

This example uses the `Companies.jmp` data table, which contains financial data for 32 companies from the pharmaceutical and computer industries.

A financial analyst wants to explore the following question:

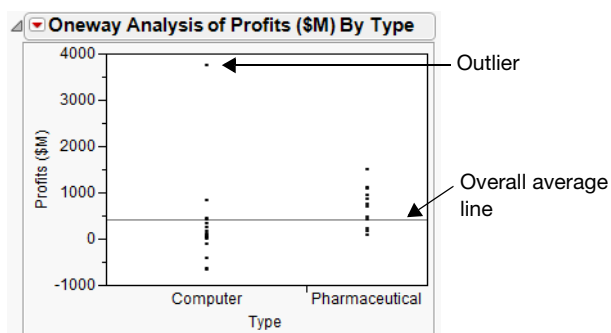
- How do the profits of computer companies compare to the profits of pharmaceutical companies?

To answer this question, fit Profits (\$M) by Type.

### Discovering the Relationship

1. Open the `Companies.jmp` sample data table.
2. If you still have the `Companies.jmp` sample data table open, 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**.
3. Select **Analyze > Fit Y by X**.
4. Select Profits (\$M) and click **Y, Response**.
5. Select Type and click **X, Factor**.
6. Click **OK**.

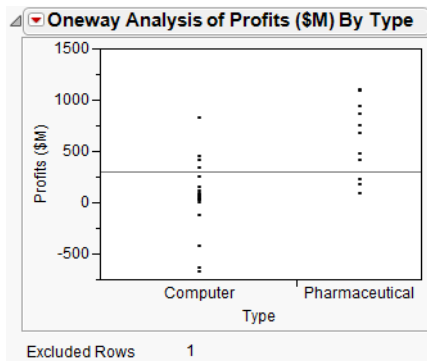
**Figure 4.15** Profits by Company Type



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 the outlier:

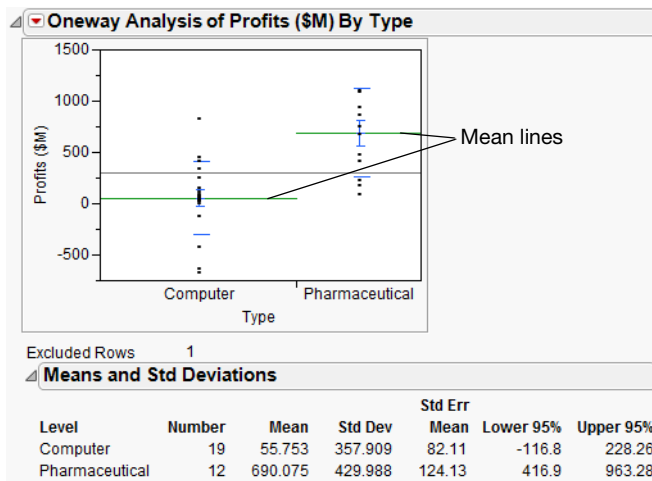
1. Click on the outlier.

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 without the outlier, select **Script > Redo Analysis** from the red triangle menu for Oneway Analysis. You can close the original Scatterplot window.

**Figure 4.16** Updated Plot


Removing the outlier gives the financial analyst a clearer picture of the data.

5. To continue analyzing the relationship, select these options from the red triangle menu for Oneway Analysis:
  - **Display Options > Mean Lines**. This adds mean lines to the scatterplot.
  - **Means and Std Dev**. This displays a report that provides averages and standard deviations.

**Figure 4.17** Mean Lines and Report


## Interpreting the Results

The financial analyst wanted to know how the profits of computer companies compared to the profits of pharmaceutical companies. The updated scatterplot shows that pharmaceutical companies have higher average profits than computer companies. In the report, if you one subtract mean value from the other, the difference in profit is about \$635 million. The plot also shows that some of the computer companies have negative profits, while all of the pharmaceutical companies have positive profits.

## Performing the t-test

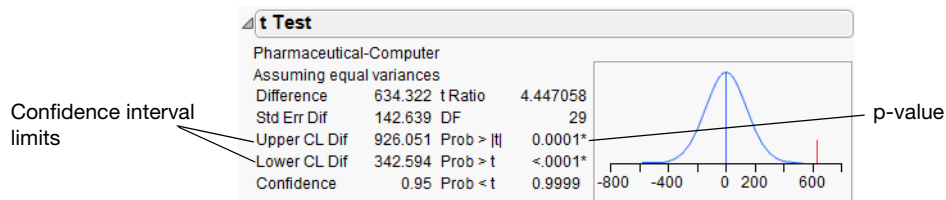
The financial analyst has looked at only a sample of companies (the companies in the data table). The financial analyst now wants to examine these questions:

- Does a difference exist in the broader population, 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, select **Means/Anova/Pooled t** from the red triangle for Oneway Analysis.

**Figure 4.18** t Test Results



The p-value of 0.0001 is less than the significance level of 0.05, which indicates statistical significance. Therefore, the financial analyst can conclude that the difference in average profits for the sample data is not due to chance alone. This means that in the larger population, the average profits for pharmaceutical companies are different from the average profits for computer companies.

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 in Figure 4.18. Using these values, the financial analyst concludes that the average profit of pharmaceutical companies is between \$343 million and \$926 million higher than the average profit of computer companies.

## Comparing Proportions

If you have categorical  $X$  and  $Y$  variables, you can compare the proportions of the levels within the  $Y$  variable to the levels within the  $X$  variable.

**Scenario**

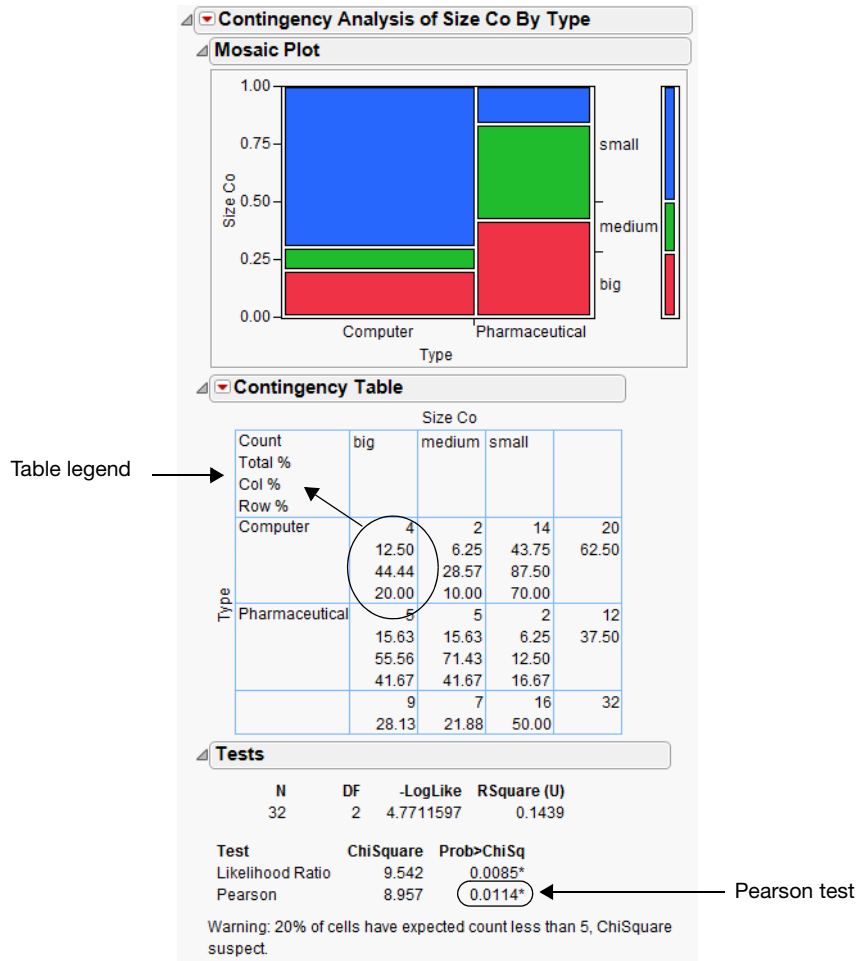
This example continues to use the `Companies.jmp` data table. In “Comparing Averages for One Variable,” p. 103, a financial analyst determined that pharmaceutical companies have higher profits on average than do computer companies.

The financial analyst wants to know whether the size of a company affects profits more for one type of company than the other? However, before examining this question, the financial analyst needs to know whether the populations of computer and pharmaceutical companies consist of the same proportions of small, medium, and big companies.

**Discovering the Relationship**

1. Open the `Companies.jmp` sample data table.
2. 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**.
3. Select **Analyze > Fit Y by X**.
4. Select `Size Co` and click **Y, Response**.
5. Select `Type` and click **X, Factor**.
6. Click **OK**.

Figure 4.19 Company Size by Company Type



The Contingency Table contains information that is not applicable for this example. From the red triangle menu for **Contingency Table** deselect **Total %** and **Col %** to remove that information. Figure 4.20 shows the updated table.

**Figure 4.20** Updated Contingency Table

		Size Co			
Count		big	medium	small	
Row %					
Type	Computer	4 20.00	2 10.00	14 70.00	20
	Pharmaceutical	5 41.67	5 41.67	2 16.67	12
		9	7	16	32

### Interpreting the Results

The statistics in the Contingency Table are graphically represented in the Mosaic Plot. Together, the Mosaic Plot and the Contingency Table compare the percentages of small, medium, and big companies between the two industries. 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.

### Interpreting the Test

The financial analyst has looked at only a sample of companies (the companies in the data table). The financial analyst needs to know whether the percentages differ in the broader populations of all computer and pharmaceutical companies.

To answer this question, use the p-value from the Pearson test in the **Tests** report. See Figure 4.19. Since the p-value of 0.011 is less than the significance level of 0.05, the financial analysts concludes the following:

- The differences in the sample data are not due to chance alone.
- The percentages differ in the broader population.

Now the financial analyst knows that the proportions of small, medium, and big companies are different, and can answer the question: Does the size of company affect profits more for one type of company than the other?

## Comparing Averages for Multiple Variables

The section “Comparing Averages for One Variable,” p. 103, compared averages across the levels of a categorical variable. To compare averages across the levels of two or more variables at once, use the *Analysis of Variance* technique (or ANOVA).

### Scenario

The financial analyst can answer the question that we started to work through in the Comparing Proportions section, which is: Does the size of the company have a larger effect on the company’s profits, based on type (pharmaceutical or computer)?

To answer this question, compare the company profits by these two variables:

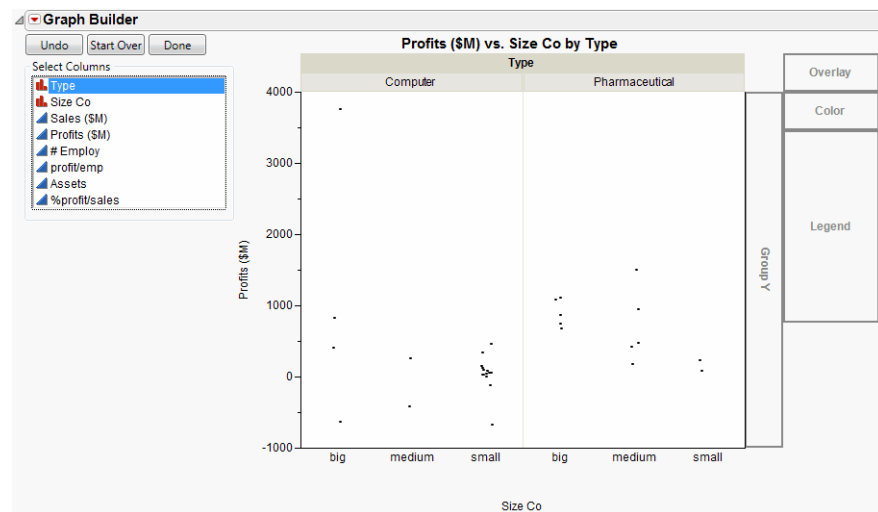
- Type (pharmaceutical or computer)
- Size (small, medium, big)

### Discovering the Relationship

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

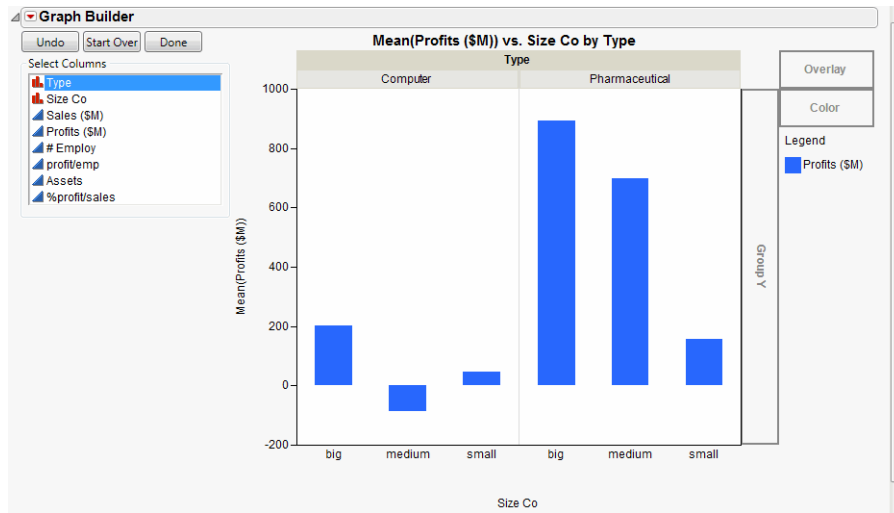
1. Open the Companies.jmp sample data table.
2. Select **Graph > Graph Builder**. The Graph Builder window appears.
3. Click on **Profits (\$M)** and drag and drop it into the **Y** zone.
4. Click on **Size Co** and drag and drop it into the **X** zone.
5. Click on **Type** and drag and drop it into the **Group X** zone.

**Figure 4.21** Graph of Company Profits



The graph shows 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.

6. 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.
7. Right-click in the graph and select **Points > Change To > Bar**. Comparing mean profits is easier with bar charts than with points.

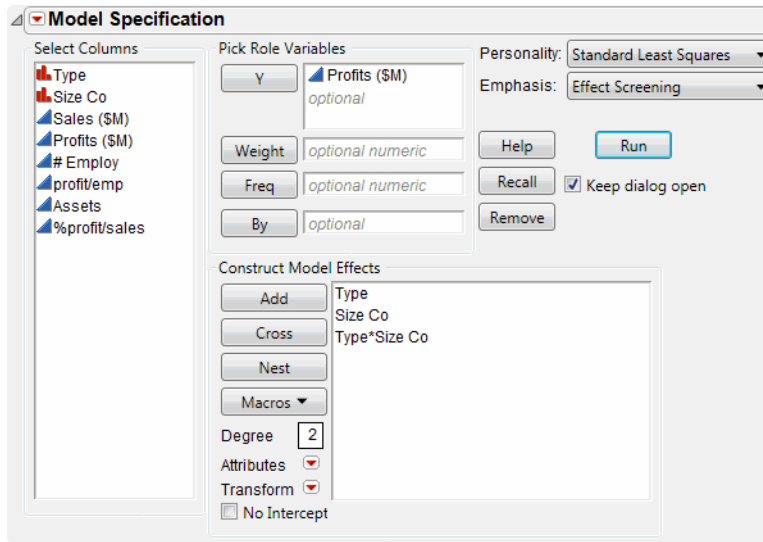
**Figure 4.22** Graph with Outlier Removed


The updated graph shows that pharmaceutical companies have higher average profits. The graph also shows that profits differ between company sizes for only the pharmaceutical companies. When the effect of one variable (company size) changes for different levels of another variable (company type), this is called an *interaction*.

### Quantifying the Relationship

Since this data is only a sample, the financial analyst needs to determine the following:

- if the differences are limited to this sample and due to chance
    - or
  - if the same patterns exist in the broader population
1. Open the Companies.jmp sample data table.
  2. Select **Analyze > Fit Model**.
  3. Select Profits (\$M) and click **Y**.
  4. Select both Type and Size Co.
  5. Click the **Macros** button and select **Full Factorial**.
  6. From the Emphasis menu, select **Effect Screening**.
  7. Select the **Keep dialog open** option.

**Figure 4.23** Completed Fit Model Window

8. Click **Run**. The report window shows the model results.

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

**Note:** For complete details about all of the **Fit Model** results, see *Modeling and Multivariate Methods*.

### Using Effect Tests

The Effect Tests report (see Figure 4.24) 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 4.24** Effect Tests Report

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Type	1	1	1,401,847	10.1368	0.0039*
Size Co	2	2	724,616.2	2.6198	0.0927
Type*Size Co	2	2	448,061.5	1.6200	0.2180

First, look at the test for the interaction in the model: the Type\*Size Co effect. Figure 4.22 showed that the pharmaceutical companies appeared to have different profits between company sizes. However, the effect test indicates that there is no interaction between type and size as it relates to profit. The p-value of 0.218 is large (greater than the significance level of 0.05). 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.

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**.

**Figure 4.25** Updated Effect Tests Report

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Type	1	1	1,356,298	9.3768	0.0049*
Size Co	2	2	434,161.3	1.5008	0.2410

The p-value for the Size Co effect is large, indicating that there are no differences based on size in the broader population. The p-value for the Type effect is small, indicating that the differences that you saw in the data between computer and pharmaceutical companies is not due to chance.

### Conclusions

The financial analyst wanted to know whether the size of the company has a larger effect on the company's profits, based on type (pharmaceutical or computer). The financial analyst can now answer the question as follows:

- There is a real difference in profits between computer and pharmaceutical companies in the broader population.
- There is no correlation between the company's size and type and its profits.

## Using Regression with Multiple Predictors

The section "Using Regression with One Predictor," p. 99 showed you how to build simple regression models consisting of one predictor variable and one response variable. *Multiple regression* predicts the average response variable using two or more predictor variables.

### Scenario

This example uses the **Candy Bars.jmp** data table, which contains nutrition information for candy bars.

A dietician wants to predict calories using the following information:

- Total fat
- Carbohydrates
- Protein

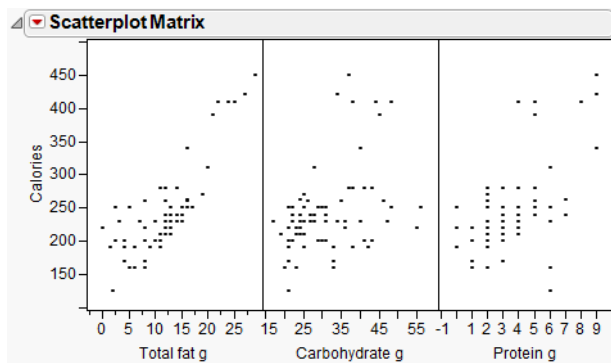
Use *multiple regression* to predict the average response variable using these three predictor variables.

### Discovering the Relationship

To visualize 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. Select **Calories** and click **Y, Columns**.
4. Select **Total fat g**, **Carbohydrate g**, and **Protein g**, and click **X**.
5. Click **OK**.

**Figure 4.26** Scatterplot Matrix Results

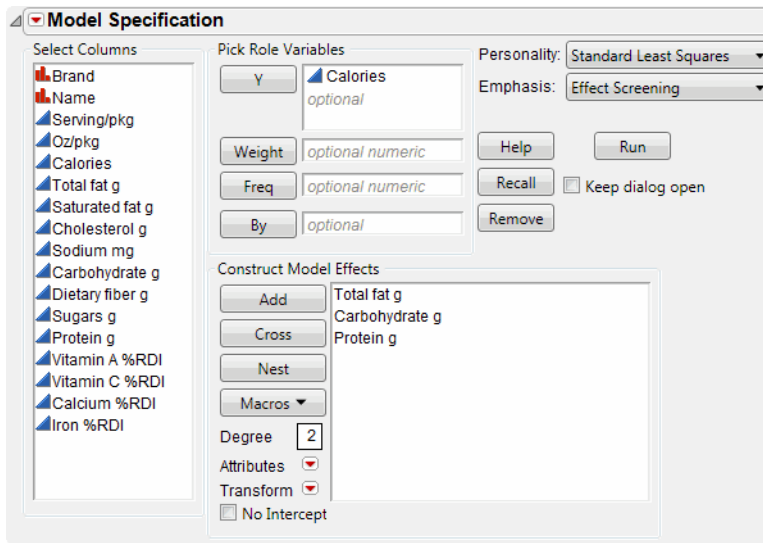


The scatterplot matrix shows that there is a positive correlation between calories and all three variables. The correlation between calories and total fat is the strongest. Now that the dietician knows that there is a relationship, the dietician can build a multiple regression model to predict average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

### Building the Multiple Regression Model

Continue to use the Candy Bars.jmp sample data table.

1. Select **Analyze > Fit Model**.
2. Select **Calories** and click **Y**.
3. Select **Total Fat g**, **Carbohydrate g**, and **Protein g** and click **Add**.
4. Next to Emphasis, select **Effect Screening**.

**Figure 4.27** Fit Model Window


5. Click **Run**.

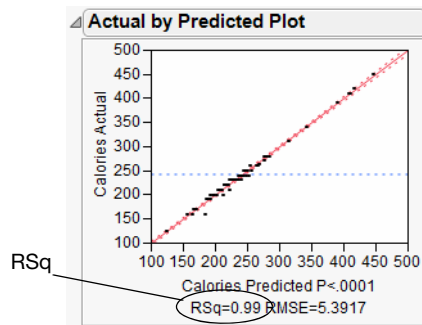
The report window shows the model results. To interpret the model results, focus on these areas:

- “Using the Actual by Predicted Plot,” p. 114
- “Interpreting the Parameter Estimates,” p. 115
- “Using the Prediction Profiler,” p. 116

**Note:** For complete details about all of the model results, see *Modeling and Multivariate Methods*.

### ***Using the Actual by Predicted Plot***

The Actual by Predicted Plot shows the actual calories versus the predicted calories. As the predicted values come closer to the actual values, the points on the scatterplot fall closer around the red line. See Figure 4.28. Because the points are all very close to the line, you can see that the model predicts calories based on the chosen factors well.

**Figure 4.28** Actual by Predicted Plot

Another measure of model accuracy is the RSq value (which appears below the plot in Figure 4.28). 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.

### Interpreting the Parameter Estimates

The Parameter Estimates report shows the following information:

- The model coefficients
- P-values for each parameter

**Figure 4.29** Parameter Estimates Report

	Model coefficients		P-values	
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-5.964301	2.899986	-2.06	0.0434*
Total fat g	8.9899516	0.144981	62.01	<.0001*
Carbohydrate g	4.097505	0.071025	57.69	<.0001*
Protein g	4.4013313	0.39785	11.06	<.0001*

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. For example, suppose that 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:

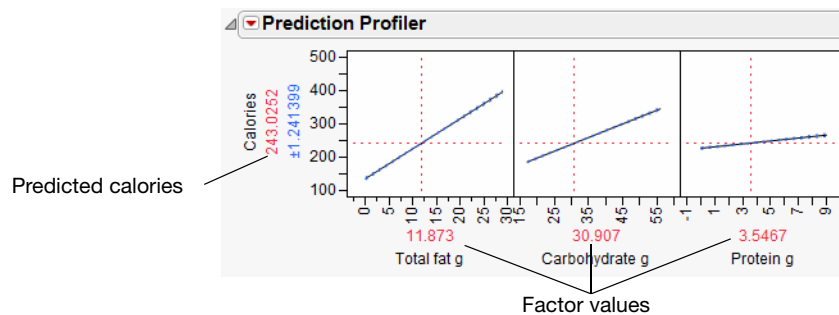
$$277.92 = -5.9643 + 8.99*11 + 4.0975*43 + 4.4013*2$$

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.

### Using the Prediction Profiler

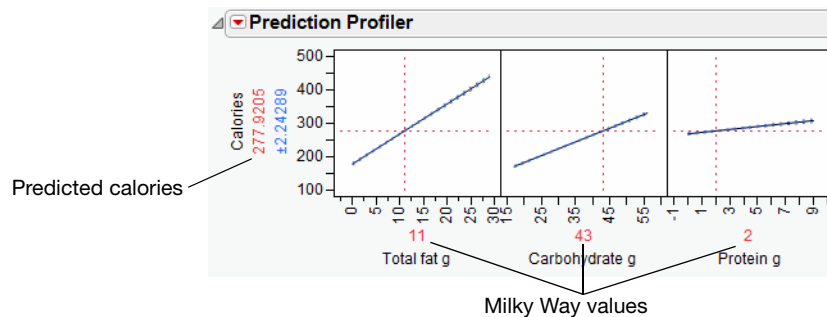
Use the Prediction Profiler to see how changes in the factors affect the predicted values. The profile lines show the magnitude of change in calories as the factor changes. The line for **Total fat g** is the steepest, meaning that changes in total fat have the largest effect on calories.

**Figure 4.30** Prediction Profiler



Click and drag the vertical line for each factor to see how the predicted value changes. You can also 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 (row 59).

**Figure 4.31** Factor Values for the Milky Way



---

**Note:** For complete details about the Prediction Profiler, see *Modeling and Multivariate Methods*.

---

### **Conclusion**

The dietician now has a good model to predict calories of a candy bar based on its total fat, carbohydrates, and protein.



## Saving and Sharing Your Work

### Saving and Recreating Results

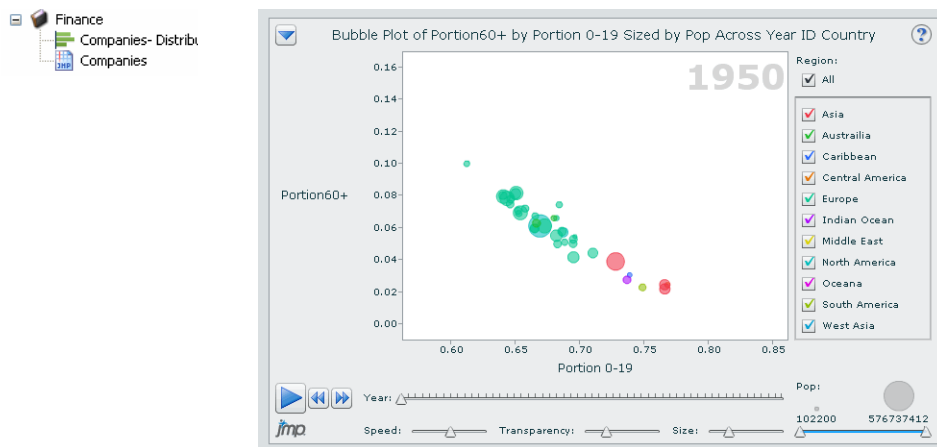
---

Once you have generated results from your data, JMP provides you with multiple ways to share your work with others. Here are some of the ways you can share your work:

- saving platform results as journals or projects
- saving results, data tables, and other files in projects
- saving scripts to reproduce results in data tables
- creating Adobe Flash Versions of platform results

---

**Figure 5.1** Examples of Saving Work



# Contents

Saving Platform Results in Journals . . . . .	121
Example: Creating a Journal . . . . .	121
Adding Additional Analyses. . . . .	122
Creating Projects . . . . .	122
Using Scripts . . . . .	123
Example: Saving and Running a Script . . . . .	123
About Scripts and JSL . . . . .	125
Creating Adobe Flash Versions of the Profiler and Bubble Plot . . . . .	125
Example: Saving an Adobe Flash Version of a Bubble Plot. . . . .	125

## Saving Platform Results in Journals

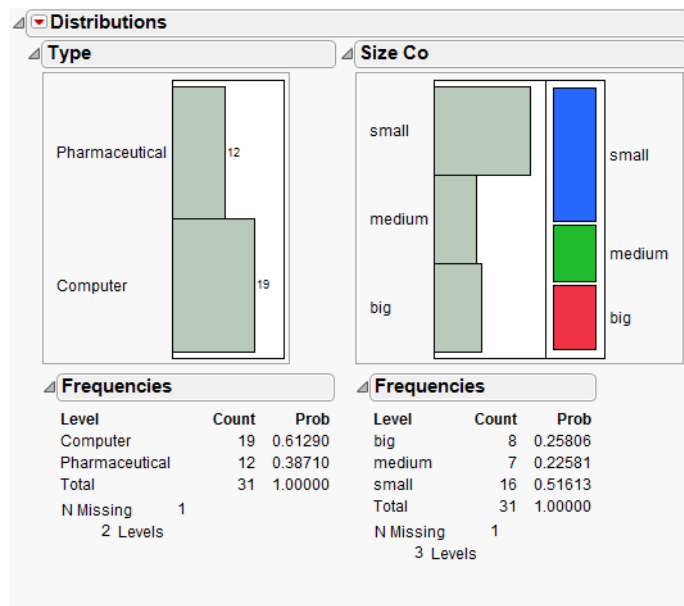
Save platform reports for future viewing by creating a journal of the report window. The journal is a copy of the report window. You can edit or append additional reports to an existing journal. The journal is not connected to the data table. A journal is an easy way to save the results from several report windows in a single report window that you can share with others.

### Example: Creating a Journal

Create and modify a report and save it to a journal

1. Open the Companies.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select both Type and Size Co and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu for **Type**, select **Histogram Options > Show Counts**.
6. From the red triangle menu for **Size Co**, select **Mosaic Plot**.
7. Select **Edit > Journal** to journal these results. 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. In the Type bar chart, if you click on the Computer bar, no rows are selected in the data table.

Since the journal is a copy of your results, most of the red triangle menus do not exist. A journal does have a red triangle menu for each new report that you add to the journal. This menu has two options:

**Rerun in new window** If you have the original data table that was used to create the original report, this option runs the analysis again. The result is a new report window.

**Edit Script** This option opens a script window that contains a JSL script to recreate the analysis. JSL is a more advanced topic that is covered in the *Scripting Guide*.

## Adding Additional Analyses

If you perform another analysis, you can add the results of the analysis to the existing journal.

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

---

## Creating Projects

Save multiple JMP file types (such as data tables, reports, journals, and scripts) in 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**

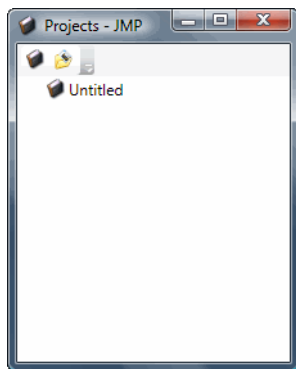
#### **Create a report**

1. Open the Companies.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Profits (\$M) and profit/emp, and click **Y, Columns**.
4. Click **OK**.

#### **Create a project and add the data table and the report to it**

1. To start a new project, select **File > New > Project**.

---

**Figure 5.3** Initial Project Window

- 
2. Right-click the project (Untitled) and select **Rename**, and then enter a new name (Finance).
  3. To add the Distribution results to the project, right-click on the project name and select **Add Window**.
  4. In the Select a Window window, select the Distribution results.
  5. Click **OK**. The Distribution results are added to the project.
  6. To add the Companies.jmp data table to the project, repeat step 3 and select the Companies data table from the window.
  7. Click **OK**. The data table is added to the project.

---

**Figure 5.4** The Final Project

---

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

---

## Using Scripts

Everything in JMP is scriptable, meaning that any action that you perform can be saved as a JMP Scripting Language (JSL) script. You can use a script to reproduce your actions or results at any time.

### Example: Saving and Running a Script

Create a report

1. Open the Companies.jmp sample data table.

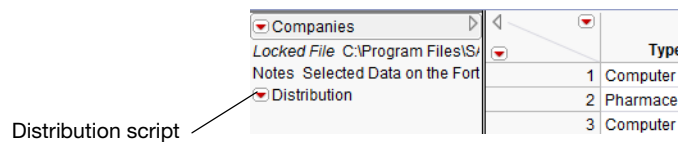
2. Select **Analyze > Distribution**.
3. Select **Type** and **profit/emp** and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu for **Type**, select these options:
  - **Histogram Options > Show Counts**
  - **Confidence Interval > 0.95**
6. From the red triangle menu for **profit/emp**, select these options:
  - **Outlier Box Plot**, to remove the outlier box plot
  - **CDF Plot**
7. From the red triangle menu for **Distributions**, select **Stack**.

#### Save the script to the data table and run it

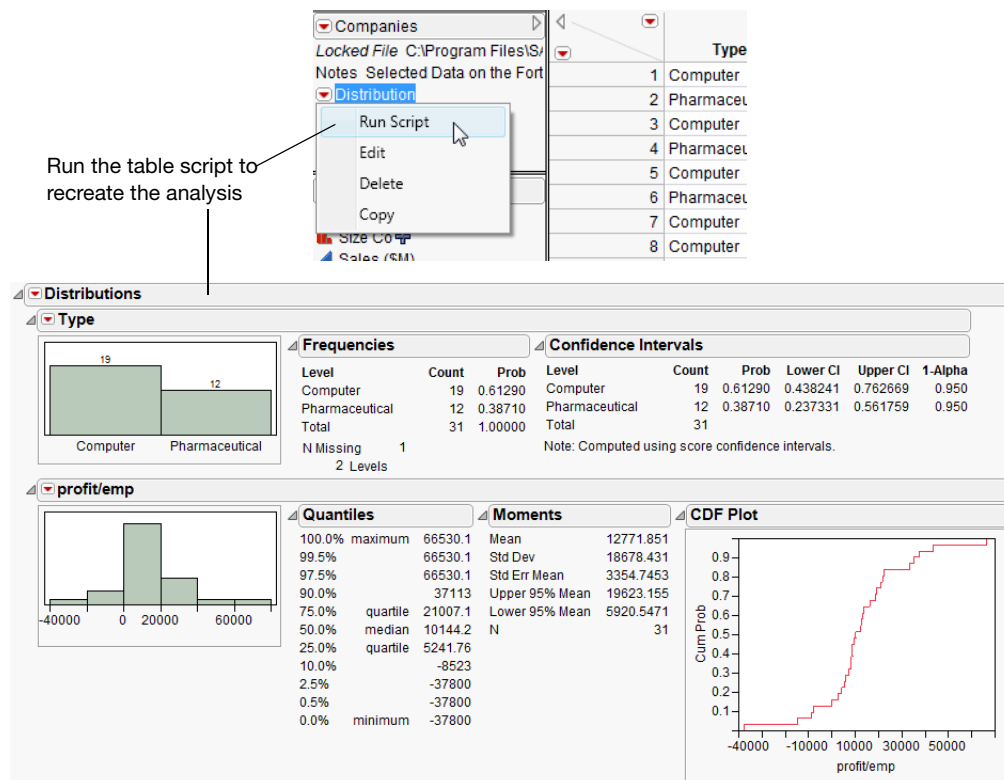
1. To save this analysis, select **Script > Save Script to Data Table** from the red triangle menu for **Distributions**. The **Distribution** script appears in the **Table** panel.

---

**Figure 5.5** Distribution Script



- 
2. Close the **Distribution** report window.
  3. To recreate the analysis, select **Run Script** from the red triangle menu for the **Distribution** script.

**Figure 5.6** Running the Distribution Script

## About Scripts and JSL

The script that you saved in this section contains JMP Scripting Language (JSL) commands. JSL is a more advanced topic that is covered in the *Scripting Guide*.

## Creating Adobe Flash Versions of the Profiler and Bubble Plot

To share the interactive results of a profiler or bubble plot outside of JMP, you can export a .SWF file and use the Adobe Flash Player to view it. You can import the .SWF file into presentations and applications. You can also save the results as an HTML page with the .SWF output embedded.

### Example: Saving an Adobe Flash Version of a Bubble Plot

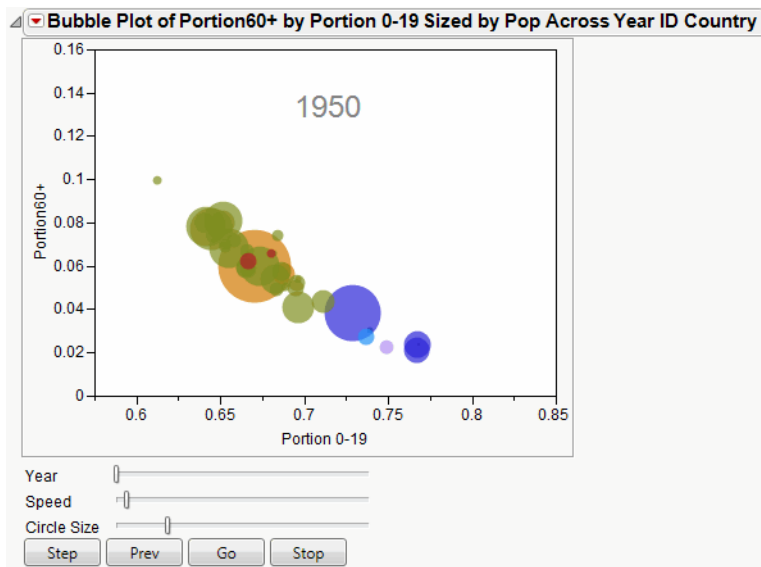
Create a bubble plot in JMP

1. Open the PopAgeGroup.jmp sample data table.

## Creating Adobe Flash Versions of the Profiler and Bubble Plot

2. Select **Graph > Bubble Plot**.
3. Select **Portion60+** and click **Y**.
4. Select **Portion 0-19** and click **X**.
5. Select **Country** and click **ID**.
6. Select **Year** and click **Time**.
7. Select **Pop** and click **Sizes**.
8. Select **Region** and click **Coloring**.
9. Click **OK**.

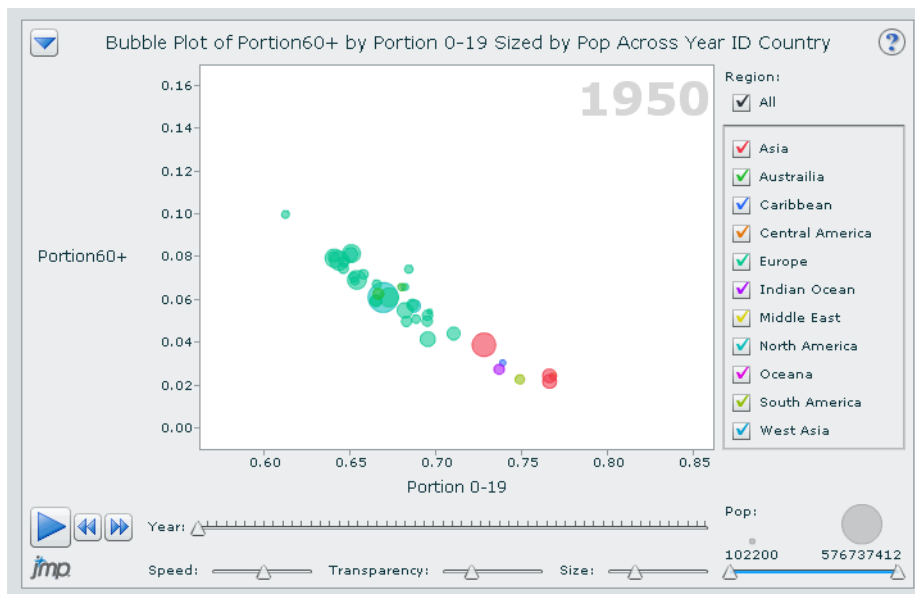
---

**Figure 5.7** Initial Bubble Plot

---

**Save the bubble plot as a .SWF file**

1. Select **Save As Flash (SWF)** from the red triangle menu.
2. In the Save As SWF window, select the location to which you want to save the file.
3. Click **Save**. The Adobe Flash version of the Bubble Plot is saved as HTML and appears in a web browser.

**Figure 5.8** Adobe Flash Version of Bubble Plot

### Additional Information

The JMP Web site has additional information about the following:

- more details about the Adobe Flash versions of the Profiler and Bubble Plot platforms
- instructions on importing the Adobe Flash versions into Microsoft PowerPoint 2007

This information is on the JMP Web site here:

[www.jmp.com/support/swfhelp/en/](http://www.jmp.com/support/swfhelp/en/)



# Chapter 6

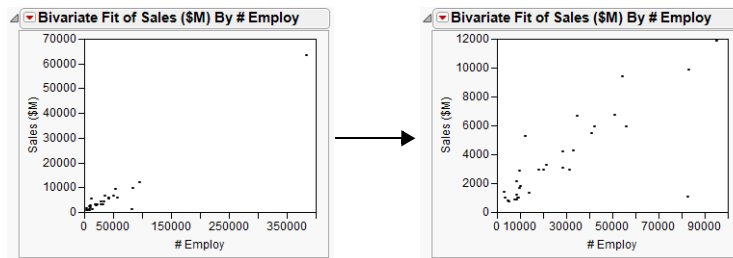
## Special Features

### Automatic Updating and Integrating with SAS

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

- Update analyses or graphs automatically
- Customize platform results
- Integrate with SAS to use advanced analytical features

**Figure 6.1** Examples of Special Features



```
Candy BarsModel - JMP
File Edit Tables DOE Analyze Graph Tools View Window Help
DATA Candy_Bars; INPUT  Calories Total_fat_g Carbohydrate_g Protein_g; Lines;
310 20 28 6
230 12 27 4
220 12 24 3
170 8 21 3
200 2.5 43 1
260 16 26 5
;
RUN;

PROC GLM DATA=Candy_Bars ALPHA=0.05;
MODEL  Calories = Total_fat_g Carbohydrate_g Protein_g;
RUN;
```

# Contents

Automatically Updating Analyses and Graphs . . . . .	131
Example: Using Automatic Recalc . . . . .	131
Changing Preferences . . . . .	134
Example: Changing Preferences . . . . .	135
Integrating JMP and SAS . . . . .	137
Example: Creating SAS Code . . . . .	138
Example: Submitting SAS Code . . . . .	138

---

## Automatically Updating Analyses and Graphs

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, include, 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 *Basic Analysis and Graphing*.
- 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. However, for Variability/Gauge Chart, Capability, and Control Chart, Automatic Recalc is turned off by default.

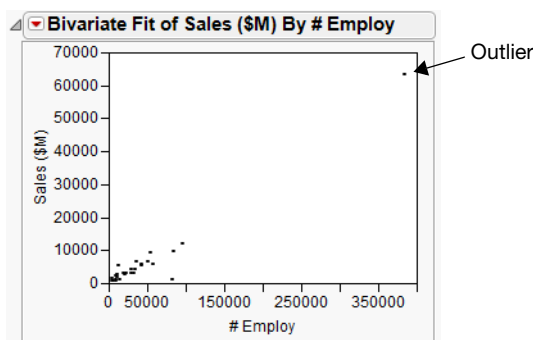
### Example: Using Automatic Recalc

This example uses the Companies.jmp sample data table, which contains 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. Select Sales (\$M) and click **Y, Response**.
4. Select # Employ and click **X, Factor**.
5. Click **OK**.

---

**Figure 6.2** Initial Scatterplot



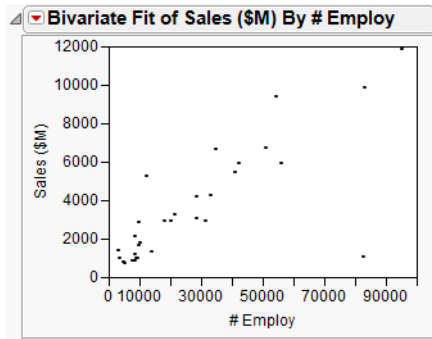
---

The initial scatterplot shows 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.

## Automatically Updating Analyses and Graphs

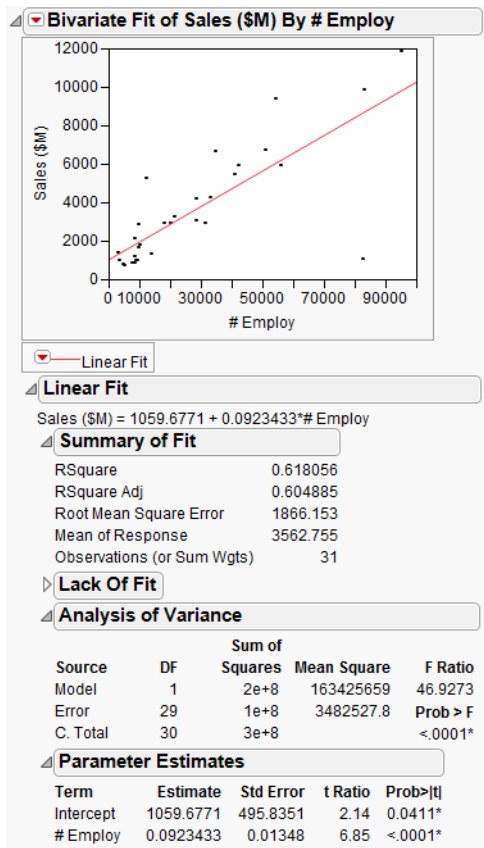
6. Turn on Automatic Recalc by selecting **Script > Automatic Recalc** from the red triangle menu.
7. Click on the outlier to select it.
8. Select **Rows > Exclude/Unexclude**. The point is excluded from the analysis and the scatterplot is automatically updated.

---

**Figure 6.3** Updated Scatterplot

If you fit a regression line to the data, the point in the lower right corner is an outlier, and influences the slope of the line. If you then exclude the outlier with Automatic Recalc turned on, you can see the slope of the line change.

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

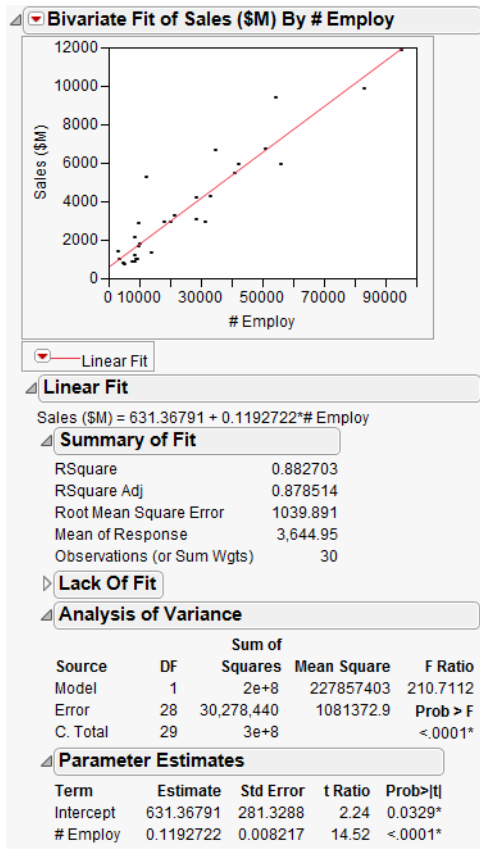
**Figure 6.4** Regression Line and Analysis Results

10. Click on the outlier to select it.

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

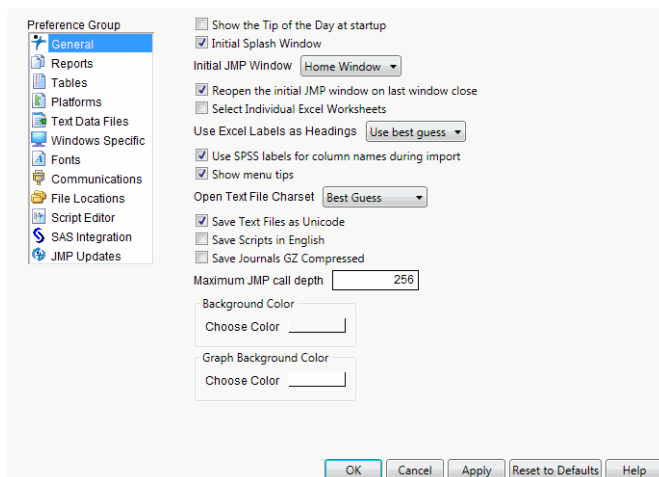
When you exclude a point, the analyses are recalculated without the data point, but the data point is not hidden in the scatterplot. To also hide the point in the scatterplot, select the point, and then select **Rows > Hide/Unhide**.

Figure 6.5 Updated Regression Line and Analysis Results



## Changing Preferences

You can change preferences in JMP using the Preferences window. To open the Preferences window, select **File > Preferences**.

**Figure 6.6** Preferences Window

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

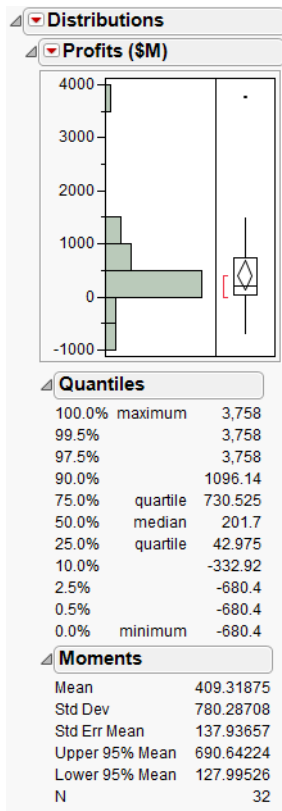
## Example: Changing Preferences

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.

This example shows how to set the Distribution platform so that an Outlier Box Plot is not added to the initial report.

### Create a Distribution using the default preference setting

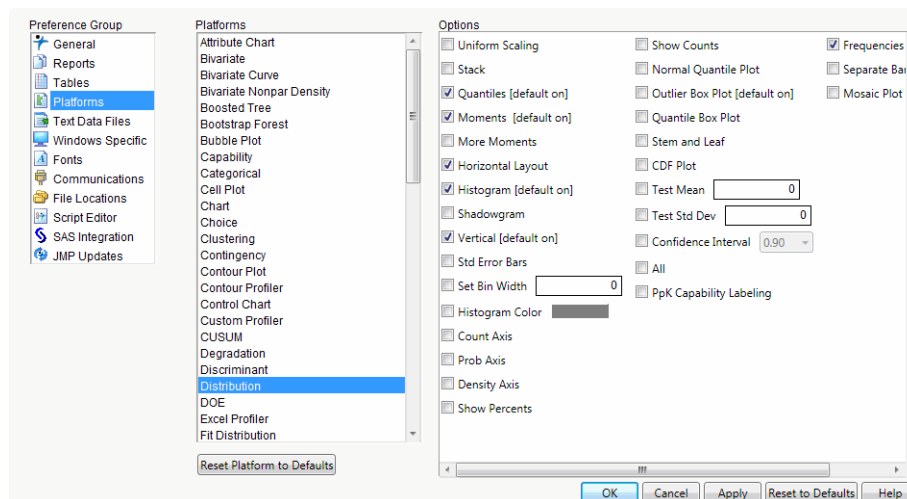
1. Open the Companies.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Profits (\$M) and click **Y, Columns**.
4. Click **OK**.

**Figure 6.7** Distribution Report Window

The histogram is vertical, and the graphs includes an outlier box plot. To change the histogram to horizontal and remove the outlier box plot, select the appropriate options from the red triangle menu for Profits (\$M). However, if you want those preferences to be in effect every time you use the platform, then change them in the Preferences window.

#### Change the preference for the Outlier Box Plot and run Distribution again

1. Select **File > Preferences**.
2. Select **Platforms** from the preference group.
3. Select **Distribution** from the Platforms list.
4. Select the **Horizontal Layout** option to turn it on.
5. Deselect the **Outlier Box Plot** option to turn it off.

**Figure 6.8** Distribution Preferences

6. Click **OK**.
7. Repeat the Distribution analysis. See “Create a Distribution using the default preference setting,” p. 135

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 *Using JMP*.

## Integrating JMP and SAS

**Note:** You must have access to SAS, either on your local machine or on a server, to use SAS through JMP.

Using JMP, you can interact with SAS as follows:

- Write or create SAS code in JMP.
- Submit SAS code and view the results in JMP.
- Connect to a SAS Metadata Server or a SAS Server on a remote machine.
- Connect to SAS on your local machine.
- Open and browse SAS data sets.
- Retrieve and view data sets generated by SAS.

For complete details about integrating JMP and SAS, see the *Using JMP*.

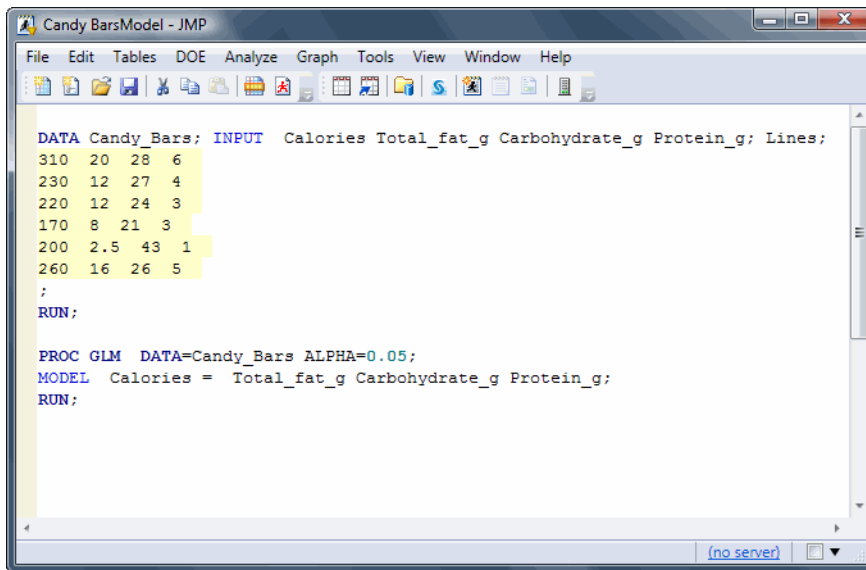
## Example: Creating SAS Code

This example uses the Candy Bars.jmp sample data table, which contains nutrition data for candy bars.

1. Open the Candy Bars.jmp sample data table.
2. Select **Analyze > Fit Model**.
3. Select **Calories** and click **Y**.
4. Select **Total fat g**, **Carbohydrate g**, and **Protein g**, and click **Add**.
5. From the red triangle menu for Model Specification, select **Create SAS Job**.

Figure 6.9 shows the SAS code. (Not all of the data is shown.)

**Figure 6.9** SAS Code

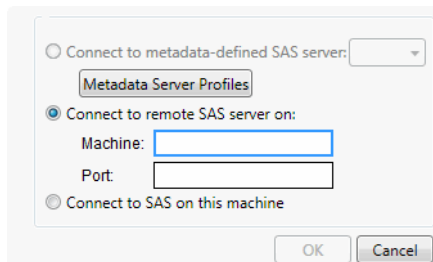


```
DATA Candy_Bars; INPUT Calories Total_fat_g Carbohydrate_g Protein_g; Lines;
310 20 28 6
230 12 27 4
220 12 24 3
170 8 21 3
200 2.5 43 1
260 16 26 5
;
RUN;

PROC GLM DATA=Candy_Bars ALPHA=0.05;
MODEL Calories = Total_fat_g Carbohydrate_g Protein_g;
RUN;
```

## Example: Submitting SAS Code

1. Open the Candy Bars.jmp sample data table.
2. Select **Analyze > Fit Model**.
3. Select **Calories** and click **Y**.
4. Select **Total fat g**, **Carbohydrate g**, and **Protein g**, and click **Add**.
5. From the red triangle menu for Model Specification, select **Submit to SAS**.
6. In the **Connect to SAS Server** window (see Figure 6.10), choose a method to connect to SAS (if you are not already connected). For this example, select **Connect to SAS on this machine**.

**Figure 6.10** Connect to SAS Server

7. Click OK.

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 6.11 shows the results formatted as a JMP report. For details, see *Using JMP*.

**Figure 6.11** SAS Results Formatted as a JMP Report

The SAS System  
The GLM Procedure

**The GLM Procedure**

**Data**

**Number of Observations**

Number of Observations Read	75
Number of Observations Used	75

Dependent Variable: Calories

**Analysis of Variance**

**Calories**

**Overall ANOVA**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	282,358	94119.3	3237.58	7.7e-76
Error	71	2064.03	29.0709		
Corrected Total	74	284,422			

**Fit Statistics**

Calories			
R-Square	Coeff Var	Root MSE	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	185,260	185,260	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	111,777	111,777	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



### A

- Actual by Predicted plot 114
- Adobe Flash
  - creating bubble plots 125
  - creating Profiler results 125
- analyses
  - adding to journals 122
  - analyzing variable relationships 98
  - automatically updating 131–134
- Anscombe, F.J. 87
- automatic recalc 131–134
- averages, comparing
  - multiple variables 108
  - single variable 103

### B

- bar charts 57
  - creating 58
  - example 57–59
  - interacting with 59
  - interpreting 58
- books, Help 24
- box plots 67
  - changing to line plots 78
  - creating 68
  - Distribution platform and 94
  - example 68–69
  - interpreting 69
  - outlier box plots 20
- bubble plots 80
  - creating 81
  - creating Adobe Flash version 125
  - example 81–84
  - interacting with 83
  - interpreting 83
  - options 83
  - trail bubbles 83

### C

- categorical variables
  - comparing averages 103–105
  - comparing proportions 105–108
  - example 95–98
  - relationships 98
- cells, selecting or deselecting 35
- character data type 38
- charts
  - bar charts 57–59
  - creating separate charts 79
  - variability charts 73–75
- code, SAS
  - creating 138
  - submitting 138
- Column Info window 37–38
- column names 38
  - changing 38
  - entering 38
  - unique 38
- Column Properties menu 39
- columns
  - adding 31
  - locking 39
  - naming 38
  - selecting or deselecting 34
- columns panel 18
- contents, Help 24
- contingency tables 107
- continuous modeling type 38, 90
- continuous variables 90
  - comparing averages 103–105
  - example of analyzing 93
  - fitting lines 100–102
  - modeling types 90
  - regression with one predictor 99–102
  - relationships 98
- copying and pasting data 27
- correlation 60

**D**

data

- changing
  - modeling types 91
  - multiple values 32
  - single value 32
- copying and pasting 27
- creating patterned 33
- deselecting
  - cells 35
  - columns 34
  - rows 34
- excluding data points 62
- filtering 41–42
- finding means 47
- graphing 87
- hiding data points 62
- importing 27
- including data points 62
- moving into JMP 27–28
- searching for values 36
- selecting
  - cells 35
  - columns 34
  - data points 64
  - rows 34
- unexcluding 62

data filter 41–42

data grid 17

data points

- excluding 62
- hiding 62
- including 62
- selecting 64

data tables 13

*See also* modeling types

adding

- columns 31
- rows 30
- to projects 123

changing name of 31

Column Properties menu 39

columns panel 18

creating 30

data grid 17

data types 38

example 25

formatting numeric values 39

joining 49

opening 27

- opening sample 16
- rows panel 18
- sorting by columns 51
- subsetting 47–49
- table panel 17
- typing data into 29–31

data types 38

- character 38
- numeric 38
- row state 38

disclosure buttons 13, 20

DisplayBox Scripting index 24

Distribution platform 93

- box plot and 94
- example, categorical variables 95–98
- example, continuous variables 93
- histograms and 94
- Linear Fit report 101
- Moments report 94
- Quantiles report 94
- rotating the report window 94
- t-tests 95

**E**

Effect Tests report 111

excluding data 62

**F**

Fill options 33

filtering data 41–42

Fit Model platform

- Actual by Predicted plot 114
- analyzing variable relationships 98
- Effect Tests report 111
- launching 113
- Parameter Estimates report 115
- Prediction Profiler 116

Fit Y by X platform

- analyzing variable relationships 98
- contingency table 107
- launching 103
- Means and Std Dev report 104
- Mosaic Plot 108
- Tests report 108
- t-tests 105

fitting lines 100–102

formula editor 39–41

**G**

- Graph Builder 75
  - changing box plots to line plots 78
  - creating graphs 76
  - creating separate charts 79
  - example 76–80
  - interpreting graphs 79
  - launching 109
- graphing data 87
- graphs
  - automatically updating 131–134
  - creating 76
  - multiple variable 59
  - See also* Graph Builder
  - single variable 55
- grid lines, adding to overlay plots 71

**H**

- Help
  - about JMP 24
  - books 24
  - contents 24
  - DisplayBox Scripting index 24
  - index 24
  - JMP user community 24
  - JMP.com 24
  - JSL Functions index 24
  - Object Scripting index 24
  - sample data 24
  - searching 24
  - statistical terms 24
  - Tip of the Day 24
  - tutorials 24
  - using 23
- Help menu 23
- histograms 55
  - creating 56
  - Distribution platform and 94
  - example 55–57
  - interacting with 57, 97
  - interpreting 56

**I**

- icons, modeling type 31
- importing data 27
- indexes
  - DisplayBox Scripting 24
  - Help 24

- JSL Functions 24
- Object Scripting 24
- Statistics 24

interaction 110–112

**J**

- JMP <sup>II</sup>
  - copying and pasting data into 27
  - creating SAS code in 138
  - getting started 13–22
  - importing data 27
  - integrating with SAS 137
  - JMP Web site 24
  - starting JMP 14
  - submitting SAS code in 138
  - tutorials 24
  - versus Microsoft Excel 23
- JMP scripting language 24, 125
- JMP Starter window 14
- JMP User Community 24
- JMP.com 24
- joining data tables 49
- journals
  - adding analyses 122
  - creating 121
  - saving platform results 121
- JSL (JMP scripting language) 24, 125
- JSL Functions index 24

**L**

- lasso tool 64
- launch windows 13
- line plots, changing box plots to 78
- linear fit report 101
- lines
  - adding mean lines 104
  - fitting 100
- locking columns 39

**M**

- mean lines 104
- mean, testing 95
- Means and Std Dev report 104
- menu, Help 23
- Microsoft Excel
  - importing data 27
  - versus JMP 23
- modeling types

*See also* categorical variables  
 changing 91  
 continuous 38, 90  
 example 90  
 icons 31  
 nominal 38, 90  
 ordinal 38, 90  
 Moments report 94  
 Mosaic Plot 108  
 multiple regression 112–113

## N

names  
   columns 38  
   data tables 31  
 nominal modeling type 38, 90  
   *See also* categorical variables  
 numeric data type 38  
 numeric values, formatting 39

## O

Object Scripting index 24  
 ordinal modeling type 38, 90  
   *See also* categorical variables  
 outlier box plots 20  
 outliers 57, 69  
   example 99, 101  
   excluding 101  
   implications of 103, 109  
   removing 62  
 overlay plots 69–70  
   adding grid lines 71  
   connecting points 71  
   example 70–73  
   interacting with 71  
   interpreting 71  
   legends 72  
   multiple variables and 72

## P

Parameter Estimates report 115  
 patterned data, creating 33  
 platforms 13, 18  
   examples 19–22  
   launch windows 13  
   launching 19  
   report windows and 13, 20–21  
   saving results 121

  viewing results 19  
 plots  
   *See specific plot type*  
 Prediction Profiler 116  
 predictor variables 98  
 preferences  
   changing 135  
 preferences, changing 134–137  
 Profiler platform  
   creating Adobe Flash results 125  
 projects  
   adding  
     data tables 123  
     results 123  
   creating 122  
 proportions, comparing 105–108

## Q

Quantiles report 94

## R

red triangle menus 13, 20  
 regression  
   multiple 112–113  
   one predictor 99–102  
   two predictors 112  
 report windows 20  
   closing or opening 22  
   description 13  
   disclosure buttons 13, 20  
   interacting with 21  
   platform options 13  
   red triangle menus 13, 20  
   reports in 13  
   rotating 94  
 reports  
   Effect Tests 111  
   Linear Fit 101  
   Means and Std Dev 104  
   Moments 94  
   Parameter Estimates 115  
   Quantiles 94  
   Tests 108  
 results, saving platform 121  
 rotating the report window 94  
 row state data type 38  
 rows  
   adding 30  
   clear row states 106

- excluding or unexcluding 62
- hiding or unhiding 62
- returning to default state 106
- selecting or deselecting 34

rows panel 18

## S

sample data tables 24

- opening 16

SAS

- creating code 138
- integrating with JMP 137
- submitting code 138

saving

- platform results 121

scatterplot matrices 64

- creating 65
- example 65–67
- interacting with 66
- interpreting 66

Scatterplot Matrix platform

- example 113
- launching 113

scatterplots 60–61

- example 61
- interacting with 63
- interpreting 62

scripts

- creating 123
- JSL and 125
- running 123

searching

- for values 36
- Help 24

side-by-side box plots

- See* box plots

sorting

- data tables by columns 51

standard deviations 104

Starter window 14

starting JMP 14

Statistics index 24

statistics, summary 43–45

subsetting

- data tables 47–49

summary statistics 43–45

## T

table panel 17

testing means 95

Tests report 108

Tip of the Day window 14, 24

trail bubbles 83

t-tests

- Distribution platform 95
- Fit Y by X platform 105
- two-sample 105

tutorials 24

## V

values

- changing multiple 32
- changing single 32
- formatting numeric 39
- searching for 36

variability 68

variability charts 73

- creating 73
- example of 73–75
- grouping order 74
- hiding 74
- interpreting 75

variables

- analyzing relationships 98
- categorical 95–98, 103–108
- comparing averages for multiple 108
- comparing averages for single 103
- continuous 90, 93, 98–105
- graphs with multiple 59
- graphs with single 55
- modeling types 90–93
- overlay plots with multiple 72
- predictor 98
- types of relationships 98

## W-Z

Web site for JMP 24

windows

- JMP Starter window 14
- Tip of the Day window 14, 24





